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

Nathan Schiff
Shanghai University of Finance and Economics

Graduate Urban Economics, Week 7
April 8, 2024

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" (*a*)

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

Uses boundary fixed effects K_b *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

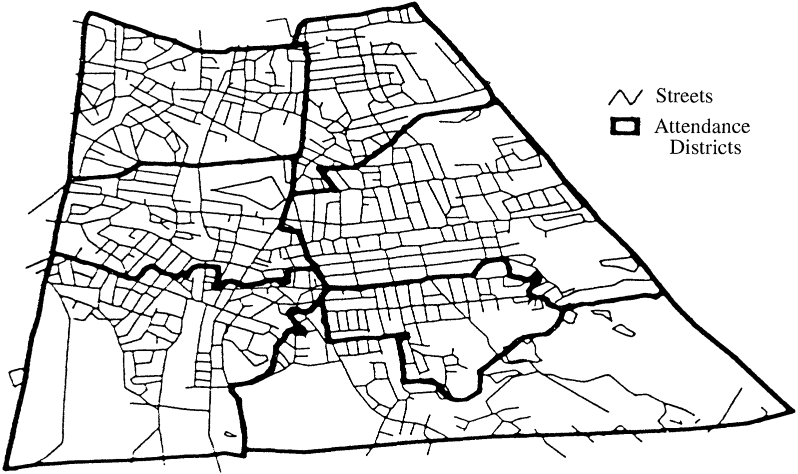


FIGURE I
Example of Data Collection for One City: Melrose
Streets, and Attendance District Boundaries

Black 1999: block groups and attendance districts

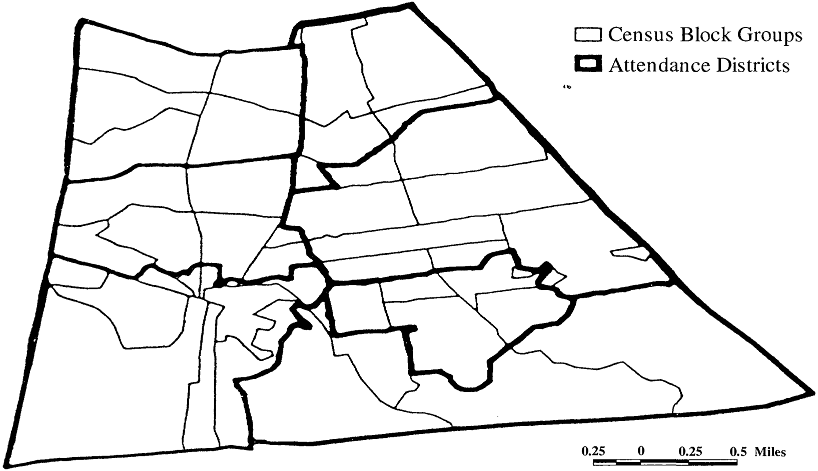


FIGURE II
Example of Data Collection for One City: Melrose
Census Block Groups and Attendance District Boundaries

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

Hedonic Estimation

and

Estimates of MWTP in Bayer Ferreira and McMillan 2007

Quick Theory of Hedonic Estimation (Taylor, 2008)

Consumer j with characteristics a^j has preferences over a house Z with n attributes and a numeraire good X: $U^j(X, z_1, z_2, ..., z_n; a^j)$

The housing market is perfectly competitive such that a house (or housing bundle) with characteristics $\bar{z} = z_1, z_2, ...z_n$ has price $P(\bar{z})$

With income y^j the consumer's budget constraint is: $y^j = X + P(\bar{z})$

Utility maximization implies: $\frac{\partial U^j}{\partial z_i} / \frac{\partial U^j}{\partial X} = \frac{\partial P}{\partial z_i} \equiv P_{zi}$

P_{zi} is the implicit price of attribute i, which is also equal to the marginal willingness to pay for a small increase in attribute i

A key assumption of this model is that each attribute is continuous so that households choose the exact level of the attribute to maximize their own utility

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

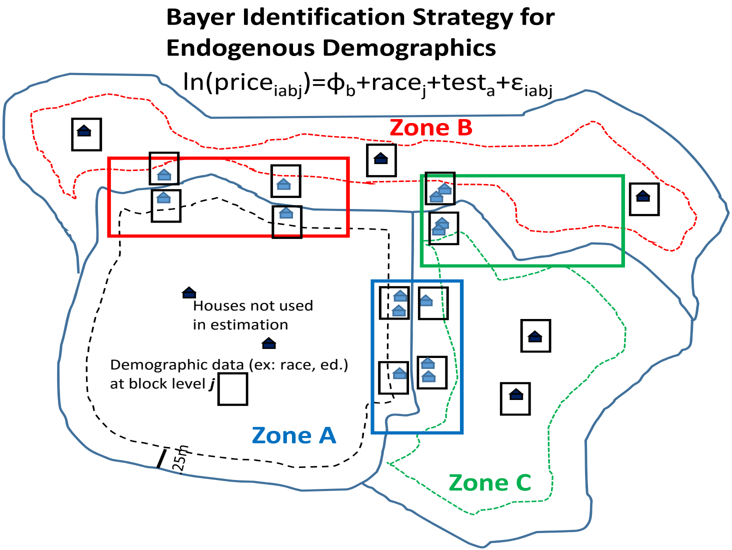
Ex: blacks in US have lower incomes and education on average than whites

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

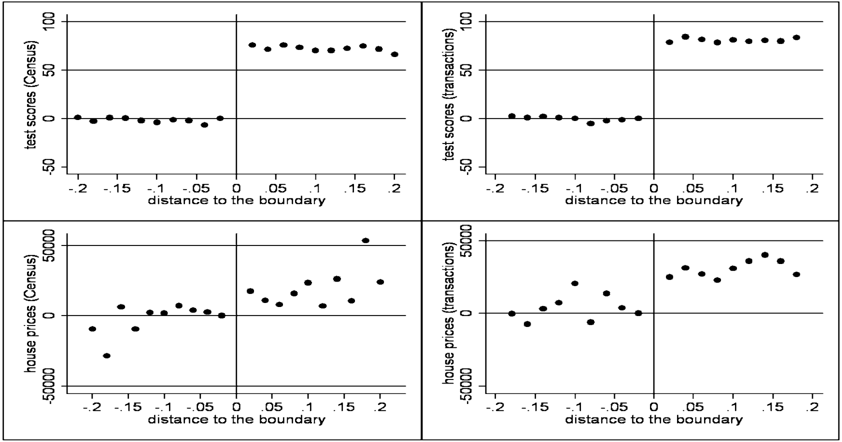
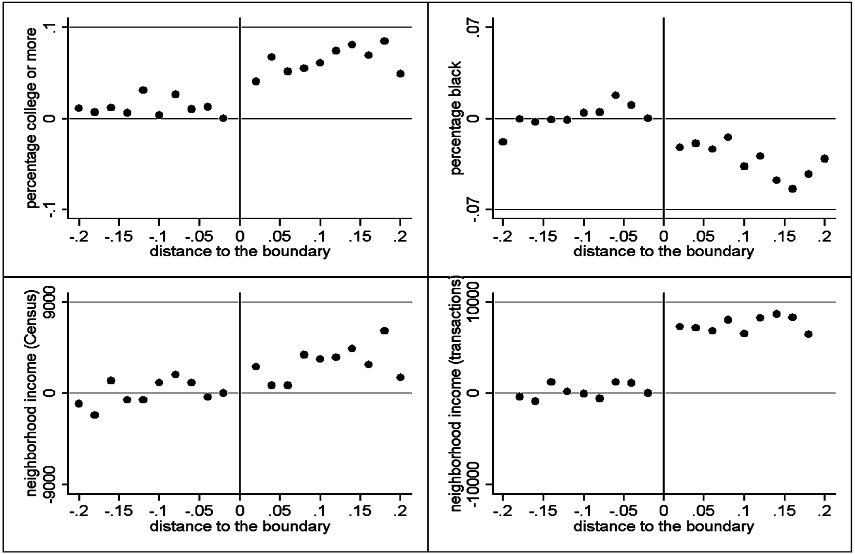


FIG. 1.—Test scores and house prices around the boundary. Each panel is constructed using the following procedure: (i) regress the variable in question on boundary fixed effects and on 0.02-mile band distance to the boundary dummy variables; (ii) plot the coefficients on these distance dummies. Thus 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 for Schools, Additional Estimates

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.

Hedonic Estimation Issues in the Presence of Sorting

Heterogeneity and MWTP

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

If consumers are heterogeneous and an attribute is discrete (few levels, ex: only 2 schools), then coefficients on a given attribute may represent MWTP of consumer who most values that attribute, not mean MWTP (first figure, next slide)

When attributes are continuously distributed (many levels, ex: many schools of different quality levels) then more consumers on the margin between two levels, thus the hedonic estimate will be closer to mean MWTP (second figure)

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

Illustration of MWTP Heterogeneity with Discrete Good

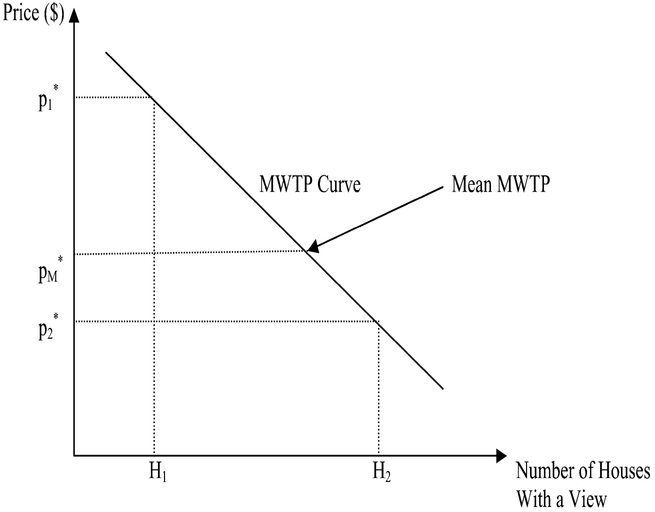


FIG. 5.—Demand for a view of the Golden Gate Bridge

MWTP Heterogeneity for Continuous Good

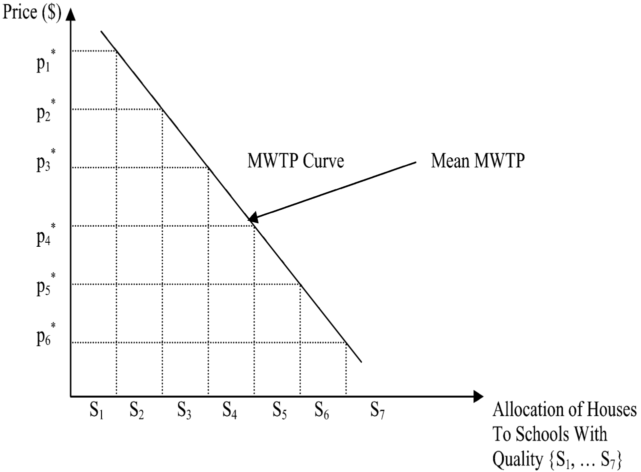


FIG. 6.—Demand for school quality

Hedonics in the Presence of Sorting

Bayer and McMillan (Hedonic Methods Handbook, Ch 10, 2008) use a simple model to demonstrate an issue with hedonic estimates in the presence of sorting

Two groups (white, black) with preferences over percentage black in neighborhood j:

$U_{ij} = \beta_i \times PctBlack_j - p_j$ with $\beta_i \sim f_b()$ for blacks, $\beta_i \sim f_w()$ for whites

Simple example: assume $f_b = U[-200, +200]$ and $f_w = U[-1500, 100]$, there are $J = 20$ neighborhoods, and 20% of population is black, 80% is white

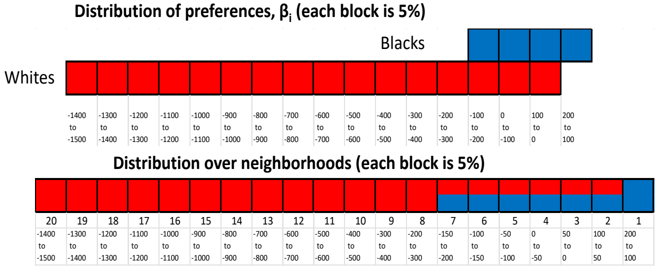
Equilibrium prices p_j in each neighborhood adjust so that marginal individual (black or white) is indifferent

Assignment to Neighborhoods in Equilibrium

To solve this simple model, order households by preferences for percentage black and then assign to neighborhoods

Since top 5% of β distribution (100-200) is only black households, first neighborhood is only black

Remaining black population preferences overlap with white preferences, with equal population for each β_i; over this range neighborhoods are integrated in equal proportions. Remaining population of city is all white.



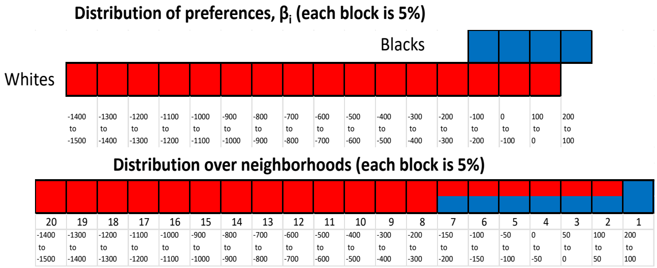
Equilibrium Prices

Marginal individual in neighborhood 2, with $\beta_i = 100$, must be indifferent between 1 and 2

$\beta_i \times PctBlack_2 - p_2 = \beta_i \times PctBlack_1 - p_1$: $100 * 0.5 - p_2 = 100 * 1 - p_1$

Condition for marginal individual in $j = 8$: $-200 * 0 - p_8 = -200 * 0.5 - p_7$

Implies $p_1 - p_2 = 50$ and $p_7 - p_8 = p_2 - p_8 = -100$; prices are only defined up to an additive constant, so normalize $p_2 = 0$:



Hedonic Estimate vs Mean Preferences

Table 10.1 Equilibrium Distribution of Neighborhood: Example 1

N'hood	% of Population	% Black	Range of β Dist	Equilibrium Price
1	5%	100%	(+200,+100)	50
2–7	30%	50%	(+100,-200)	0
8–20	65%	0%	(-200,-1,500)	100

Regressing p_j on $PctBlack_j$ yields a coefficient of -129, implying the difference in price between a completely white neighborhood and completely black neighborhood is -$129

But we know that mean β_i for whites is -$700; for population is -$560 (note: small mistake in handbook article for these numbers)–what accounts for the large error in the estimate?

Problem: prices determined by marginal individual and there are only three neighborhood types, thus infra-marginal preferences don't affect prices

Estimating Preferences Using a Sorting Model in Bayer, Ferreira, and McMillan, 2007

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

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

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:

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

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 eq (5) it can yield a hedonic that gives *mean* MWTP

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

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

Notice that if consumers are homogeneous then δ_h is constant for all *h*; this implies that eq (10) is just a simple hedonic

How to estimate first step?

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

Basic Procedure:

1. make arbitrary guess for all δ_h (all $\delta_h = 0$)
2. estimate λ_h^i terms with MLE; this is a logit model where variables are interaction terms
3. given estimates of λ_h^i, estimate δ_h using contraction mapping; the mapping is $\delta_h^{t+1} = \delta_h^t - ln(\sum_i \hat{P}_h^i)$
4. Re-estimate λ_h^i terms, then new vector of δ_h
5. Repeat process until finding a stable δ_h

Second step

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

In second step, authors regress δ_h estimates on covariates

Question: why bother with two step estimation? Why not just estimate interaction parameters only and then take mean coefficients to find average WTP?

Answer: by separating into two steps we can deal with endogeneity using IV; using instruments directly in a logit model is very difficult

Where is the endogeneity in eq (5)?

Identification

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

School quality, neighborhood demographics, and housing price may all be endogenous

As discussed earlier, school quality may be positively correlated with unobserved neighborhood quality; same with demographic characteristics

Authors assume that boundary fixed effects and demographic controls are sufficient to control for endogeneity in X_h (remember assumption is that demographic sorting occurs because of observable test scores)

Lastly, if the model is correct, housing price *must* be endogenous: higher values of ξ_h increase utility of house h and raise price

Instrumenting for Housing Price

Basic idea: use "competing products" as instruments

IO example: instrument for price of a model of car using a measure of how many close competitors there are to that model

Idea: more competitors should lower price of car (relevance) but do not affect utility of owning that car (exclusion restriction)

In BFM: instrument for price of house *h* using variables that describe housing characteristics more than three miles away (characteristics of neighboring houses could affect utility directly)

Then use their model to strengthen instrument (next slide)

Instrumenting for Housing Price, part 2

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

Distant houses gives an estimate for α_{0p}, they then take this estimate and predict market clearing house prices with only exogenous characteristics of houses (ex: age) and neighborhoods (lakes, topography)

$$P_h^i = \frac{exp(\beta * X_h^{exog} - \alpha_{0p} p_h)}{\sum_k exp(\beta * X_k^{exog} - \alpha_{0p} p_k)} \tag{NS1}$$

$$p_h^{t+1} = p_h^t + ln(\sum_i \hat{P}_h^i) \tag{NS2}$$

Cool idea, see NBER paper for details

Sorting Model Results

Discussion of Average Willingness to Pay Results

Find that average willingness to pay for school quality estimated using sorting model is very close to marginal willingness to pay coefficient from basic hedonic

Authors argue that this is because school quality is widely distributed (i.e., earlier figure on MWTP for continuously distributed good)

However, find that estimates for average willingness to pay for black neighbors is substantially more negative than hedonic estimates

Interpret this as racial preferences (discrimination) of non-marginal white residents (live in mostly white neighborhoods); MWTP is picking up residents who live in mixed neighborhoods and have different preferences

Estimates of Average Willingness to Pay

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	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)

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.

Discussion of Heterogeneity

Lastly, authors look at MWTP by different groups (different $\alpha * z_k$ estimates)

Find lots of sorting preferences

Find that educated households prefer to live with other educated households (pay additional $32 per month); less-educated prefer to live with other less-educated (required additional $26 to live with more educated)

Similar results by race

Estimates of Heterogeneity in MWTP

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)

NOTE.—The first row of the table reports the mean marginal willingness to pay for the change reported in the column heading. The remaining rows report the difference in willingness to pay associated with the change listed in the row heading, holding all other factors equal. The full heterogeneous choice model includes 135 interactions between nine household characteristics and 15 housing and neighborhood characteristics. The included household characteristics are household income, the presence of children under 18, and the race/ethnicity (Asian, black, Hispanic, white), educational attainment (some college, college degree or more), work status, and age of the household head. The housing and neighborhood characteristics are the monthly user cost of housing, distance to work, average test score, whether the house is owner-occupied, number of rooms, year built (1980s, 1960–79, pre-1960), elevation, population density, crime, and the racial composition (% Asian, % black, % Hispanic, % white) and average education (% college degree) and household income for the corresponding census block group. Standard errors are reported in parentheses.

References

Bayer, Patrick, McMillan, Robert, and Reuben, Kim "An Equilibrium Model of Sorting in an Urban Housing Market" *NBER Working Paper 10865*, 2004

Bayer, Patrick, Ferreira, Fernando, and McMillan, Robert, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods" *Journal of Political Economy*, 2007

Bayer, Patrick and McMillan, Robert, "Distinguishing Racial Preferences in the Housing Market: Theory and Evidence," Chapter 10 in *Hedonic Methods in Housing Markets*, 2008

Bayer, Patrick, and Timmins, Christopher, "On the equilibrium properties of locational sorting models," *Journal of Urban Economics*, 2005.

Bayer, Patrick, and Timmins, Christopher, "Estimating Equilibrium Models of Sorting across Locations," *Economic Journal*, 2007.

References continued

Kuminoff, Nicolai V., Smith, V. Kerry, and Timmings, Christopher, “The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets,” *Journal of Economic Literature*, 2013

Taylor, Laura, “Developments in Hedonic Modeling,” Chapter 1 in *Hedonic Methods in Housing Markets*, 2008