UNIVERSITY OF AMSTERDAM

AMSTERDAM BUSINESS SCHOOL Business

Heterogeneous Households and Market Segmentation in a Hedonic Framework

ASSA-AREUEA 2020
San Diego

Martijn I. Dröes (University of Amsterdam)
Steven C. Bourassa (Florida Atlantic University)
Martin E. Hoesli (University of Geneva)

Price of a heterogeneous good

???

- Price based on the characteristics of a good: $P = f(X)$.
- Reduced form equation as laid down by Rosen (1974).
- Household characteristics no longer play a direct role.

Our paper

- Since then household information has been used to:
  - Bourassa et al. (1999): Capture unobserved amenities.
  - Ekeland et al. (2004): Identify housing demand/preferences.
  - Harding et al. (2003): Analyze bargaining power.
- Our paper: household information to help define market segments. Explore Rosen's quote in more detail:

"*A clear consequence of the model is that there are natural tendencies towards market segmentation ... segmented by distinct income and taste groups ...*" (Rosen, 1974, p.40)

AMSTERDAM BUSINESS SCHOOL Business

Our contribution

- Our contribution is twofold:

  1) Redefine the hedonic price function to allow for secondhand markets using an Edgeworth box.

  - Allows us to focus on household heterogeneity only.
  - Multiple consumers, connect multiple Edgeworth boxes (trade chains) and money as intermediary good.
  - A consumer can be a buyer of some housing attributes, but a seller of others.

  - If households sort themselves into particular types of houses, then marginal prices and quantities are clustered (market segments):

'The hedonic price function is no longer continuous or unique.'

AMSTERDAM BUSINESS SCHOOL Business

Our contribution

2) Three empirical approaches that incorporate both information on household and housing characteristics.

- Interaction effects (exogenous class model).
- Unsupervised machine learning model (k-mean clustering, endogenous classes).
- Latent class model/finite mixture approach (endogenous classes).

- AHS metropolitan public use file for Louisville MSA 2013.
  - Possible to estimate these models using single wave + decent amount of observations. (Miami + location controls + ethnicity)
  - Household income and family structure (presence of children) as clustering variables.

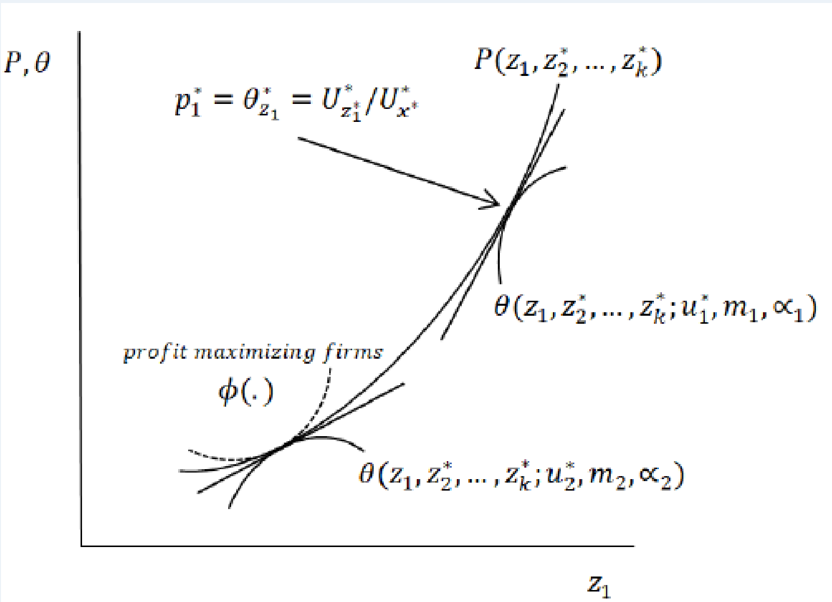
AMSTERDAM BUSINESS SCHOOL Business

Louisville

- Louisville is the 45th largest MSA.

AMSTERDAM BUSINESS SCHOOL Business

Theory: Edgeworth box



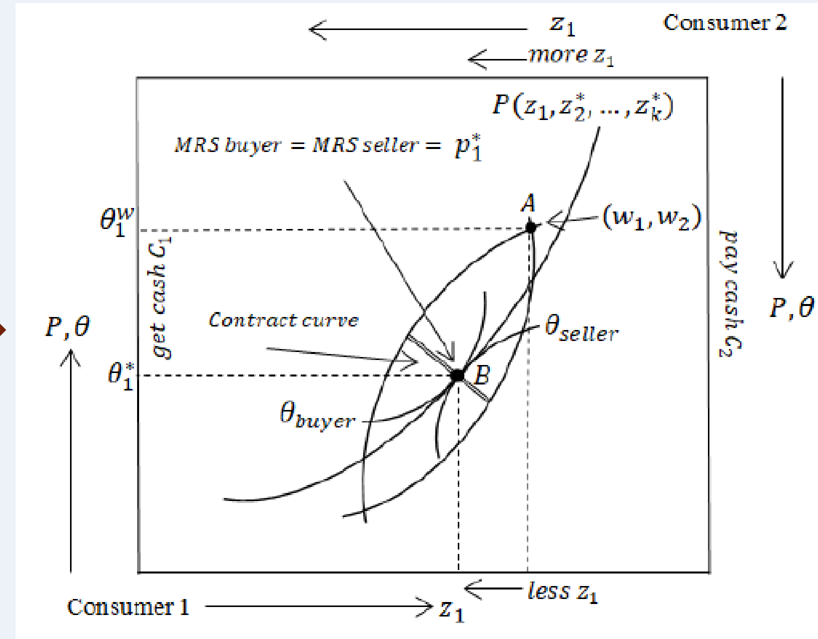


FIGURE 2—BUYERS AND SELLERS IN A SECONDHAND MARKET

- From Rosen (1974) to a secondhand market (Edgeworth box).
  - Households j are willing to pay $\theta(z; u_j, m_j, \propto_j)$ for house characteristics z given their income m_j and preferences $\propto_j$. They buy a house at the hedonic price line $P(z^*)$.
  - Edgeworth box: From endowment point A to equilibrium B, consumer 1 consumes less of z_1 and gets cash C_1 from consumer 2, either through perfect competition (Rosen, 1974) or bargaining (Harding et al. 2003).

Theory: Market segmentation

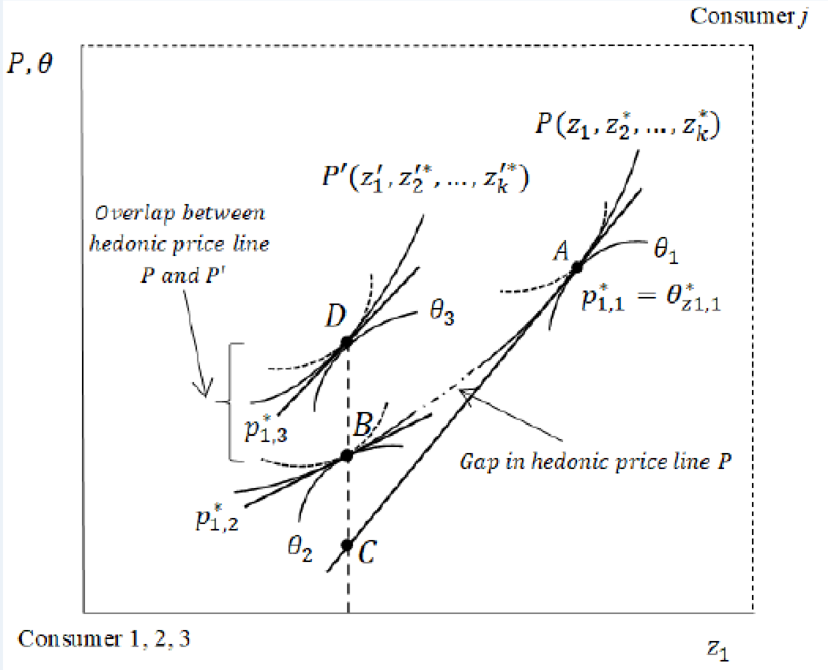


FIGURE 4—MARKET SEGMENTATION: THREE TYPES OF CONSUMERS

- Sorting of households let them trade at different parts of the hedonic price line, A vs B. Or price lines overlap, B vs D.
- We are agnostic about why such differences persist (e.g. quality differences, housing market frictions).
- Need methodology: clustering marginal price and quantities.

AMSTERDAM BUSINESS SCHOOL Business

Methodology I

- To measure differences in marginal prices:

1) Interaction effects between housing/household char.

$$\log(P_j) = \sum_j \sum_k \beta_{k,j} z_{k,j} + \varepsilon_j,$$

-easy to use, but need strong theoretical guidance.-

2) Unsupervised machine learning (k-means clustering)

$$\arg\min_{\boldsymbol{C}} \sum_{j=1}^{J} \sum_{\boldsymbol{d} \in C_j} \left\| \boldsymbol{d} - \boldsymbol{\mu}_j \right\|^2.$$

-automated, but black box.-

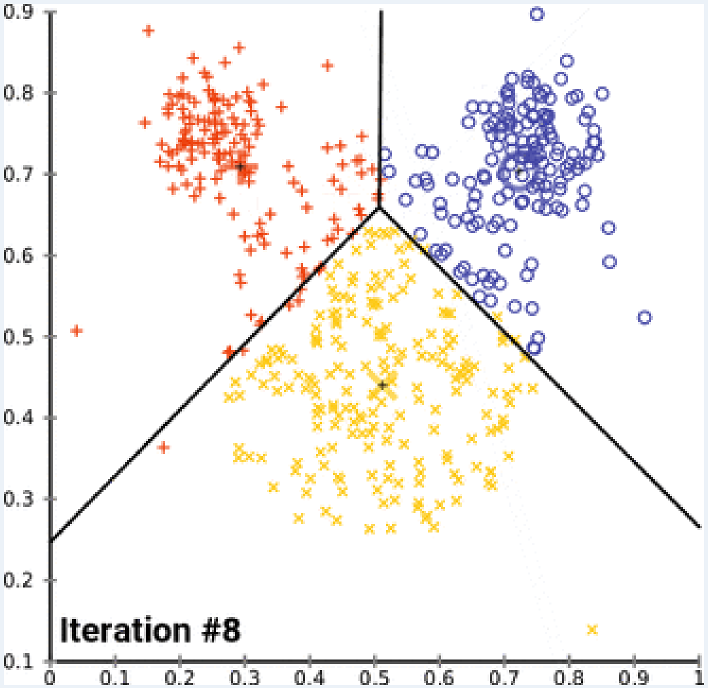
3) Full-fledged statistical approach: latent class modeling

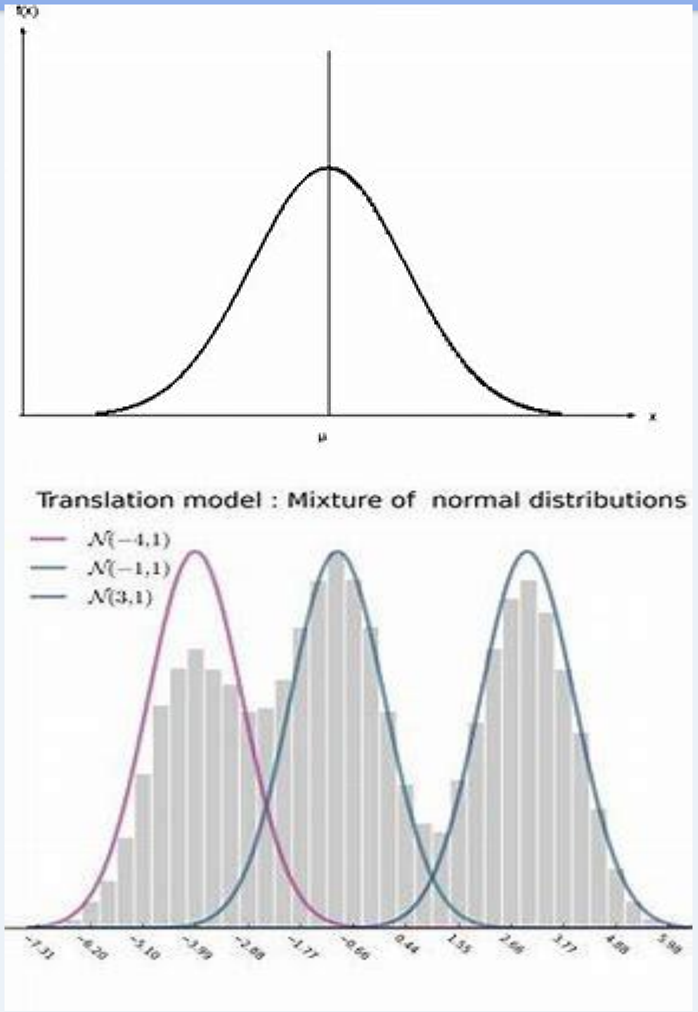
$$g(\log(P_i) \mid \boldsymbol{\mu}, \boldsymbol{\pi}) = \sum_j \pi_j^{d_{ij}} f_j\left(\log(P_i) \mid \mu_j\right)^{d_{ij}},$$

-clear about hedonic and class assignment model, standard hypothesis testing possible, but scalability is an issue.-

AMSTERDAM BUSINESS SCHOOL Business

Methodology II





Clustering
(then hedonic model)

Latent class
(interested in E(y|x) per class)

AMSTERDAM BUSINESS SCHOOL Business

Methodology III

- To measure whether there are gaps or overlaps in the distribution of trades:
- Bhattacharyya (1943) coefficient: overlap in discrete distributions:

$$BC = \sum_m q_m l_m,$$

  - m partitions, q_m and l_m proportion of members of each distribution that are part of the partition.
  - Between 0 and 1, where 1=perfect overlap.
  - Popular in pattern recognition, not often used in economics.

AMSTERDAM BUSINESS SCHOOL Business

Data

TABLE 1—SUMMARY STATISTICS: HOUSE PRICES, HOUSE CHARACTERISTICS AND HOUSEHOLD CHARACTERISTICS, LOUISVILLE (2013)

Variables	Mean	Std. Dev.	Min.	Max.
Housing variables				
Sale price (expected, $)	196,125	147,843	10,000	1,120,000
House size (sq. ft.)	2,212	1,334	99	7,235
Lot size (sq. ft.)	72,678	182,894	1	956,923
Age of structure (years)	40	24	0	94
Number of bathrooms	2.30	1.02	1	8
Number of rooms	6.64	1.76	2	13
Garage	0.79	0.40	0	1
Dishwasher	0.83	0.38	0	1
Fireplace	0.51	0.50	0	1
Floor	0.02	0.22	0	3
Louisville (former city)	0.17	0.38	0	1
Clustering variables				
Children	0.31	0.46	0	1
Household income ($)	80,319	62,546	1	456,869
Number of observations		1,636		

Note: Based on the AHS Louisville KY-IN metropolitan area public use file for 2013. Floor is the number of floors from the building main entrance to the unit, which is defined as zero for single-family houses and condominiums on the same floor as the main entrance. Children is a dummy variable for the presence of children under 18 in the household.

- We use the (log) expected sale price as dependent variable.
- For interaction effect: below/above med. inc.

AMSTERDAM BUSINESS SCHOOL Business

Results I

TABLE 2—HEDONIC MODEL AND EXOGENOUS CLASSES, LOUISVILLE (2013)

(Dependent variable: log sale price)

	(1)	(2)			
	Hedonic	Exogenous classes			F-stat.
		Reference category	Interaction children	Interaction high income	Ref. + child = Ref. + income
House size (log)	0.309***	0.251***	-0.113*	0.184***	10.72***
	(0.0383)	(0.0601)	(0.0616)	(0.0664)	
Lot size (log)	0.0185***	0.0192***	0.00506	-0.000620	
	(0.00423)	(0.00643)	(0.00921)	(0.00796)	
Age of structure	-0.00675***	-0.00557**	0.00169	-0.00481	
	(0.00156)	(0.00273)	(0.00291)	(0.00310)	
Age of structure sq.	5.63e-05***	1.93e-05	-9.65e-06	9.69e-05***	4.04**
	(1.77e-05)	(2.84e-05)	(3.37e-05)	(3.53e-05)	
Number of bathrooms	0.167***	0.138***	0.0429	0.0194	
	(0.0154)	(0.0284)	(0.0294)	(0.0308)	
Number of rooms	0.0414***	0.0457***	0.0120	-0.0225	
	(0.00815)	(0.0153)	(0.0152)	(0.0172)	
Garage	0.131***	0.148***	-0.0359	-0.0111	
	(0.0258)	(0.0420)	(0.0461)	(0.0489)	
Dishwasher	0.278***	0.303***	-0.0825	-0.0829	
	(0.0319)	(0.0444)	(0.0597)	(0.0608)	
Fireplace	0.122***	0.114***	0.0840**	-0.0185	
	(0.0222)	(0.0354)	(0.0413)	(0.0438)	
Floor	0.0347	-0.0448	0.00148	0.296**	
	(0.0695)	(0.0692)	(0.123)	(0.123)	
Louisville (former city)	0.0330	0.00749	0.163**	-0.0367	
	(0.0377)	(0.0492)	(0.0813)	(0.0817)	
Joint sig. (F-stat.)			1.82**	5.12***	1.95**
Adj. R-squared	0.637		0.652		
Observations	1,636		1,636		

Note: Robust standard errors in parentheses. High income is defined as income above the sample median of $61,000. The exogenous class model also includes children and high income as separate variables. *, **, *** indicate 10%, 5%, 1% significance, respectively.

- Interaction effect model: not so much differences.

AMSTERDAM BUSINESS SCHOOL Business

Results II

TABLE 4 —HEDONIC MODEL, CLASSES BASED ON CLUSTERING ALGORITHM, LOUISVILLE (2013)
(Dependent variable: log sale price)

	(4)						
	Three-cluster model			F-stat.			
	Cluster1	Cluster 2	Cluster 3	1 = 2	1 = 3	2 = 3	1=2=3
House size (log)	0.203***	0.267***	0.272***				8.73***
	(0.076)	(0.040)	(0.074)				
Lot size (log)	0.0188*	0.0189***	0.00897				
	(0.010)	(0.0042)	(0.011)				
Age of structure	-0.00286	-0.0102***	0.00106			7.64***	
	(0.0048)	(0.0020)	(0.0035)				
Age of structure sq.	0.0000250	0.0000884***	-0.00000628			3.50*	
	(0.000041)	(0.000030)	(0.000041)				
Number of bathrooms	0.148***	0.116***	0.194***			6.17**	10.03***
	(0.041)	(0.018)	(0.026)				
Number of rooms	0.0612***	0.0238***	-0.00326				
	(0.022)	(0.0091)	(0.017)				
Garage	0.107**	0.172***	-0.0321				
	(0.042)	(0.030)	(0.27)				
Dishwasher	0.314***	0.231**	0.0488				
	(0.042)	(0.11)	(0.044)				
Fireplace	0.117*	0.120***	0.0257				
	(0.066)	(0.022)	(0.073)				
Floor	0.0623	-0.0617	-				
	(0.091)	(0.099)					
Louisville (former city)	-0.0726	0.112*	0.204**	5.74**	7.03***		10.29***
	(0.046)	(0.062)	(0.093)				
Equality coef. (χ^2)		64.79***					
Adj. R-squared (per eq.)	0.252	0.343	0.341				
Adj. R-squared (overall)		0.662					
Observations	542	826	267				

- Bit more differences…joint classes based on income and having children.

Results III

TABLE 6 —LATENT CLASS HEDONIC MODEL, LOUISVILLE (2013)

(Dependent variable: log sale price)

	(6)						
	Three-class model			χ^2			
Hedonic variables	Class1	Class 2	Class 3	1 = 2	1 = 3	2 = 3	1=2=3
House size (log)	0.335	0.169***	0.411***			7.42***	8.38**
	(0.220)	(0.0653)	(0.0486)				
Lot size (log)	0.0649**	0.00702	0.0148*				
	(0.0316)	(0.0131)	(0.00842)				
Age of structure	-0.00714	-0.00436	-0.00995***				
	(0.0116)	(0.00313)	(0.00215)				
Age of structure sq.	7.82e-06	-3.52e-06	0.000122***			9.31***	12.29***
	(0.000114)	(4.61e-05)	(2.51e-05)				
Number of bathrooms	0.280***	0.123***	0.154***				
	(0.0911)	(0.0305)	(0.0201)				
Number of rooms	0.0188	0.0460***	0.0362***				
	(0.0511)	(0.0106)	(0.0114)				
Garage	0.121	0.0936**	0.132***				
	(0.169)	(0.0374)	(0.0329)				
Dishwasher	0.510***	0.200***	0.0999**		5.06***		7.07***
	(0.184)	(0.0629)	(0.0432)				
Fireplace	-0.0126	0.136***	0.153***				
	(0.137)	(0.0407)	(0.0298)				
Floor	0.0934	-0.157***	0.159**			10.45***	10.73***
	(0.174)	(0.0572)	(0.0690)				
Louisville (former city)	-0.109	-0.0596	0.131***			9.85***	9.91***
	(0.201)	(0.0383)	(0.0466)				
Multinomial logit variables							
Children		1.353**	0.476				
		(0.573)	(0.621)				
Household income (log)		0.125	1.183***				
		(0.113)	(0.301)				
Log pseudo likelihood		-364.32					
AIC (single class = 1,405)		819					
Adj. R-squared		0.782					
Average posterior prob.		0.716					
Entropy		0.428					
Equality coef. (χ^2)		211.31***					
Frequency, most likely class	109	637	890				
	(6.7%)	(38.9%)	(54.4%)				
Observations		1,636					

- Highly statistically significant differences.

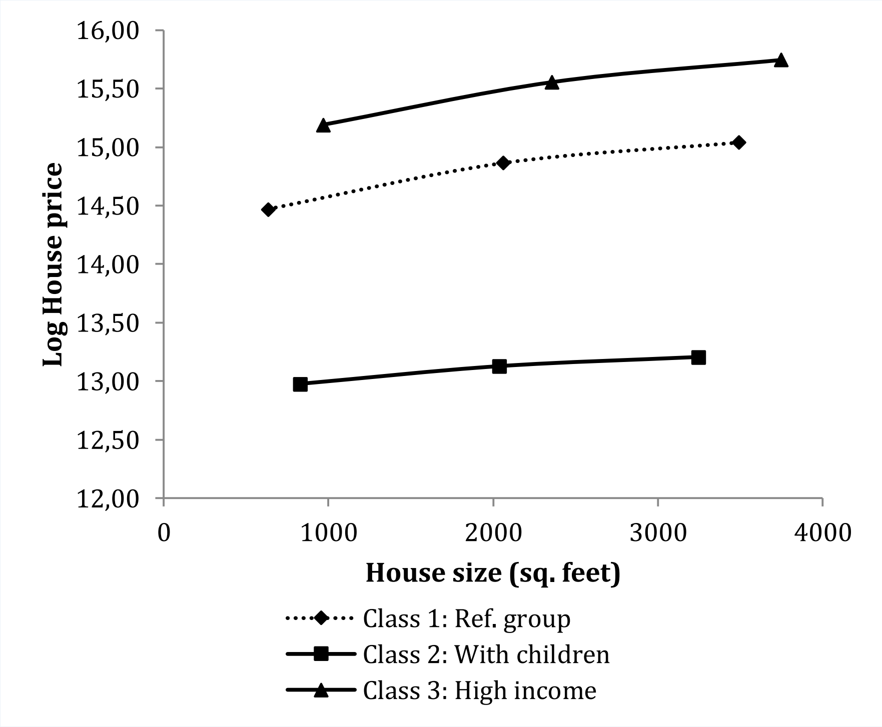
Results III – cont.

- Having children increases the probability to belong to class 2 instead of 1 (increase log odds ratio by 1.353).
- An increase in income increases the probability to belong to class 3 (increase log odds ration by 1.183).
- Seperate classes based on income and having children.
- Class 2: 45% children, Class 3: income $103,287.
- Example difference in coefficients, Floor you live on (proxy for apartment, not sig. in hedonic model):

  Class 2: discount of 15.7%, Class 3: premium of 15.9%.
- Av. Battacharyya Coefficient: 0.965 => overlaps.

AMSTERDAM BUSINESS SCHOOL Business

Segmented markets



- Hedonic price line of house size based on 3-class latent class model.

AMSTERDAM BUSINESS SCHOOL Business

Conclusion

- Household information + hedonic model to define market segments.
  - Theoretical: Edgeworth box + heterogeneous households.
  - Empirical: -3 approaches to measure differences in average marginal prices and quantities consumed.
    -Bhattacharyya coefficient (1943) to measure overlap in classes.
- Latent class seems to work best in our particular case.
- Evidence of market segmentation (overlapping price lines)
- WIP: Miami, adding more locational controls, ethnicity.

AMSTERDAM BUSINESS SCHOOL Business

Thank you for listening!

Amsterdam Business School,
Finance Department,
Real Estate Group,

m.i.droes@uva.nl