11.433J / 15.021J Real Estate Economics
Fall 2008

Real Estate Economics: Housing Attributes & Density

Sep 23, 2008

Jinhua Zhao

Outline

- Simple Richardian model expansion
  - How to price location → how to price capital
- Housing Attributes & Density
  - Housing attributes: structure, neighborhood, regional
  - Marginal utility and diminishing marginal utility
  - Expenditure vs. price
  - Why do we need hedonic model?
- Hedonic Regression Analysis:
  - Regressions: linear, log linear, log log
  - Logic of applying HRA
  - Hedonic housing price model: variation within a city (in contrast to price variation among cities)
- Land value maximization
  - First order derivatives
  - Think on the margin

Summary: Price vs. Rent; Land vs. House

	House	Land
Rent	1 $R(d) = r_a * q + c + k(b-d)$	2 $r(d) = r_a + k(b-d)/q$
Price	3 $P_t(d) = \frac{r_a q}{i} + \frac{c}{i} + \frac{k(b_t - d)}{i} + \frac{k b_t g}{i(i-g)}$	4 $p_t(d) = \frac{r_a}{i} + \frac{k(b_t - d)}{iq} + \frac{k b_t g}{i(i-g)q}$

Price accounts for future; Rent does not!

Expansion of the Ricardian model

Assumptions in a stylized city

- Monocentric: all opportunities are in the center
- Location purely defined by transportation
- houses identical: no physical differences except location
- q is fixed: 1/q density; no substitution between structure capital and land
- households identical: same income, same preference

$$\begin{cases} R(d) = r_a * q + c + k(b - d) & \textbf{Structure} \\ N * q = \pi * b^2 * V & \textbf{Boundary condition} \end{cases}$$

Expansion of the Ricardian model: relaxation of the assumptions

- Identical households → different household segments
- Identical houses → different density
- Identical houses → different characteristics
- Mono-center → multi-center

Diminishing marginal utility

Marginal Utility (MU)

Additional satisfaction obtained from consuming one additional unit of good. We might write MU as $\Delta U/\Delta x$. Graphically, MU is the slope of the utility function. In mathematical terms: dU/dx

Diminishing Marginal Utility

Principle that as more of a good is consumed, the consumption of additional amounts will yield smaller addition to utility.

The more, the merrier, but less so.

Graphs of utility and marginal utility

Hedonic prices

Hedonic Prices

- Implicit prices of attributes of differentiated goods
- Derived by observing the joint variation of product prices and bundles of product characteristics
- Expenditure vs. price

Construction of Hedonic Price Model

OLS Regression Model

$P = f(Z)$

– P = product price

– Z = vector of product attributes: structural and locational

Estimated OLS coefficients represent the **shadow prices** of product attributes (i.e., the value of an additional unit of attribute i, holding all other attributes constant)

Elements of a regression equation

Linear relationship

$$y = \alpha + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \varepsilon$$

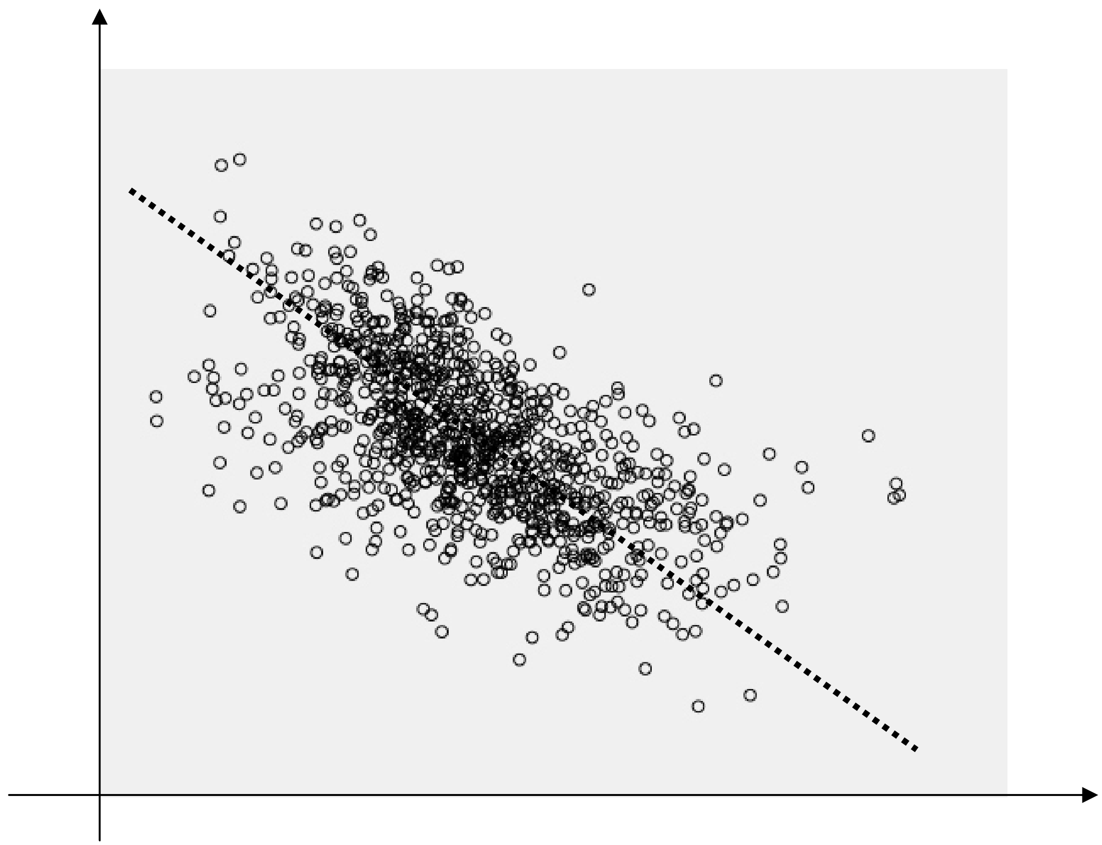
x_1, x_2, x_3	**Independent (explanatory) variables**	**Observed from data**
y	**Dependent (response) variables**	
α	**Constant term**	**Unobserved and to be estimated**
$\beta_1, \beta_2, \beta_3$	**Coefficients of the independent variables**	
ε	**Error term**	**Unobserved, Assumptions about it**

Correspondence between a linear regress and a linear

$$y = \alpha + \beta_1 * x_1 + \varepsilon$$

α **Intercept**

β **Slope**



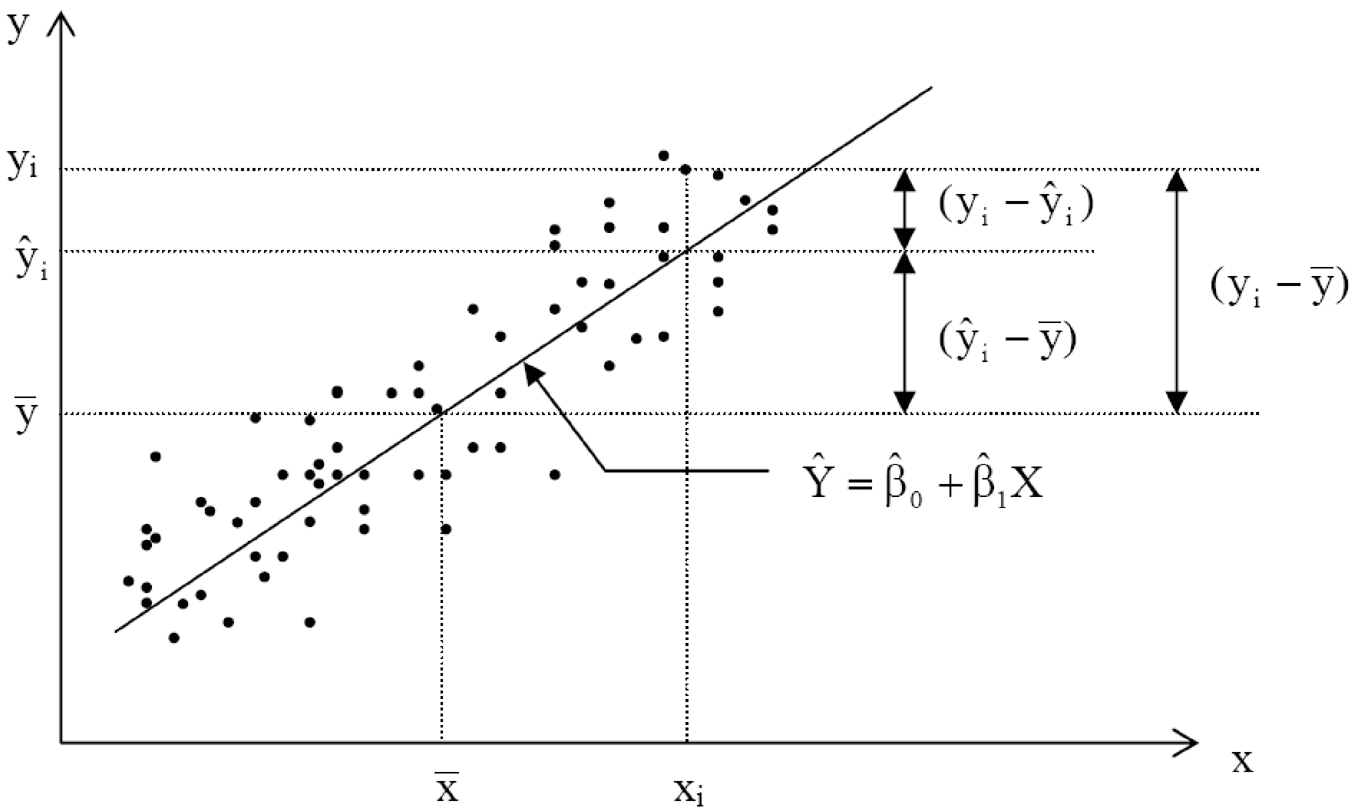
What is the best estimate of the location of the line?

- **How to define the best?**
- **How to identify it?**

Curve fitting

How to identify the best line? To minimize the distance between actual y and the estimated y

The line of best fit is the one that minimizes the sum of the squared errors



$$SSE = \sum_i (Y_i - \widehat{Y})^2$$

$$SST = \sum_i (Y_i - \overline{Y})^2$$

In order to minimize the Sum of Squared Errors $SSE = \sum_i (Y_i - \widehat{Y})^2$ **, what is the best** α **and** β

OLS (Ordinary Least Square) method

- Derive alpha and beta from the first order condition of minimizing the SSE

$$\hat{\beta} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})(X_i - \bar{X})}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$R^2 = \frac{1 - SSE}{SST}$$

Measures of goodness of fit
Standard Error of the Estimate
Coefficient of Determination: R-square: [0, 1]: % of the variance in y explained by the variances of the x
Standard error of the slope: a measure for the accuracy of betas

Statistical property of regression and assumptions of the error term

Assumptions

- 1. Regarding the shape of the relationship: linear
- 2. Regarding the expected value of the error term: 0
- 3. Regarding the variance of the error term: constant
- 4. Regarding the relationship between the error terms: independent
- 5. That the error term is normally distributed

Properties

- Unbiased: the mean of the parameter estimates is equal to the true value of the parameter that we are trying to estimate
- Efficient / best : the minimum variance of the unbiased parameter estimates

Gauss-Markov Theorem

- Assumptions 1 and 2 ➔ unbiased
- Assumptions 1, 2 and 3, 4 ➔ unbiased and efficient: best linear unbiased estimator (BLUE)
- Assumptions 1~5 ➔ useful t-statistic value

Regression: housing price variation among cities (p56)

- A simple model to explain the housing price variation among cities

- Three key factors:
  - Size of the city
  - Growth of the city
  - Construction cost

- Data:
  - 1990, CMSAs in the US

- Variables:
  - Price: median house price in 1990 (PRICE)
  - Size of the city: # of households (HH)
  - Growth of the city: % difference between 1980 and 1990 households (HHGRO)
  - Construction cost: 1990 Construction Cost Index (COST)

- Model:

$$PRICE = \alpha + \beta_1 * HH + \beta_2 * HHGRO + \beta_3 * COST + \varepsilon$$

- Expected results:
  - Size of the city
  - Growth of the city
  - Construction cost

Regression: Housing price variation among cities

- Results:

$$PRICE = -298{,}138 + 0.019 * HH + 152{,}156 * HHGRO + 1622 * COST$$

$$\begin{array}{cccc} (10.0) & (2.4) & (2.3) & (4.2) \end{array}$$

R-square=0.76

- Interpretation
  - Betas
    - Constant
    - Size of the city
    - Growth of the city
    - Construction cost
  - t-statistics
  - R2

- Notes:
  - Different scale of the variables → different scale of the betas
  - 3 variables but quite a powerful explanations
  - CMSA as the unit
  - HHGRO as the growth rate proxy

Hedonic housing price model: price variation within the city (p69)

- Expenditure vs. price / a true measure of price

- Factors
  - Number of bedrooms
  - Number of bathrooms
  - Age of structure
  - Single family attached
  - Garage
  - Poor-quality unit
  - Poor neighborhood
  - Central city

- Hedonic model

$$PRICE = \alpha + \beta_1 * BEDRMS + \beta_2 * BATHRMS + \beta_3 * GARAGE$$
$$+ \beta_4 * AGE + \beta_5 * SFA + \beta_6 * POORQUAL$$
$$+ \beta_7 * BADAREA + \beta_8 * CENTRALCITY + \varepsilon$$

- Expectations

Hedonic housing price model: price variation within the city (p69)

- Results

$$PRICE = 61508 + 13935 * BEDRMS + 50678 * BATHRMS + 21681 * GARAGE$$
$$- 60 * AGE - 3880 * SFA - 3425 * POORQUAL$$
$$- 6175 * BADAREA - 4997 * CENTRALCITY$$

R-square=0.38   N =1168   t-statistics all significant

- Interpretation

Dummy variable: a way to use nominal (qualitative) data in the regression equation

- We create one or more variables, each of which takes on the values of 0 or 1 only.
- The number of dummy variables we need is equal to k-1, where k is the number of categories in your original nominal variable.
- The regression coefficient for your dummy variable can be interpreted as the predicted change in Y when an observation is a member of the particular category, as compared to the *reference* category (explained shortly).

How NOT to use dummy variables:

Let RACE =
1 if African American
2 if Asian American
3 if Caucasian
4 if Hispanic
5 if "Other"

The correct way is to use a setof indicator ("dummy") variables and code them in this manner:
Let AFRAMER =1 if African American and 0 otherwise
Let ASIAMER =1 if Asian American and 0 otherwise
Let CAUCAS = 1 if Caucasian and 0 otherwise
Let HISPAN =1 if Hispanic and 0 otherwise
Let OTHER = 1 if "Other"and 0 otherwise
You only include 4 of the dummy variables in the regression

How do you interpret the intercept?
The intercept is the mean of the omitted group.

How do you interpret the beta of the dummy variables?
The b1coefficient is the mean of the Asiamergroup minus the mean of the Aframergroup.

Ex. Gender discrimination in the labor market

- Factors that determine earnings
  - Occupation, age, experience, education, motivation, innate ability / Race, gender (any discrimination, of concern to lawyers, planners)
- Measurement / approximation of each factor
  - year of schools as proxy for education
  - Year of working as proxy for experience

$$Earnings = \alpha + \beta_1 * School + \beta_2 * Experience + \beta_3 * Aptitude + \beta_4 * Gender + \varepsilon$$

Variables	Estimated value	Standard Error	T-statistic
Constant	4784	945	5.06
School	1146	72	15.91
Experience	285	6.8	5.74
Aptitude	39	20.2	14.13
Gender	-1867	350.5	-5.32
R-Square	0.964		

Hedonic Regression Analysis

1). Linear Hedonic equation:

X’s are structural, location attributes
no diminishing marginal utility

$$P = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

2). Log Log: (decreasing marginal utility)

A 1% change in x is associated with a β% change in P. So β is the elasticity of P with respect to x

$$P = \alpha X_1^{\beta_1} X_2^{\beta_2} \ldots\ldots X_n^{\beta_n}$$

$$\log P = \log \alpha + \beta_1 \log X_1 + \beta_2 \log X_2 + \ldots.\beta_n \log X_n$$

Logic in applying hedonic model to assess housing price

Step 1: observe other cases	**Price, Housing Attributes**
Step 2: identify pattern	**Price = f (housing attributes)**
Step 3: house in question	**Attributes of the house in question**
Step 4: estimate the price	**Apply attributes in step 3 to the function in step 2 to estimate the price of the house in question**

Underlying this process is the assumption that the “pattern”, i.e, the function f remains the same across cases.

Systematic vs. random part of a regression equation

$$y = \alpha + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \ldots + \varepsilon$$

Fact/Reality	Systematic part	Random part
	Knowledge	Ignorance
	Theory/Model	Error
	Information	Noise

1. Relationship is never perfect: error term
2. Factors contributing to the error term:
   - Measurement errors in x and/or y
   - Equation misspecification
   - Omitted variables / True relationship other than linear
   - Inherent randomness
3. R2: describe the strength of this relationship, level of knowledge, power of the theory, degrees of uncertainty

Land value maximization

House price (per floor area) and density: $P = \alpha - \beta F$

α = all housing and location factors besides FAR
F = FAR (floor area per land area)
β = marginal impact of FAR on price per square foot

Construction cost (per floor area) and density: $C = \mu + \tau F$

μ = "baseline" cost of construction
τ = marginal impact of FAR on cost per square foot

Question: what if construction technology improves or consumers' preferences change?

Land value maximization

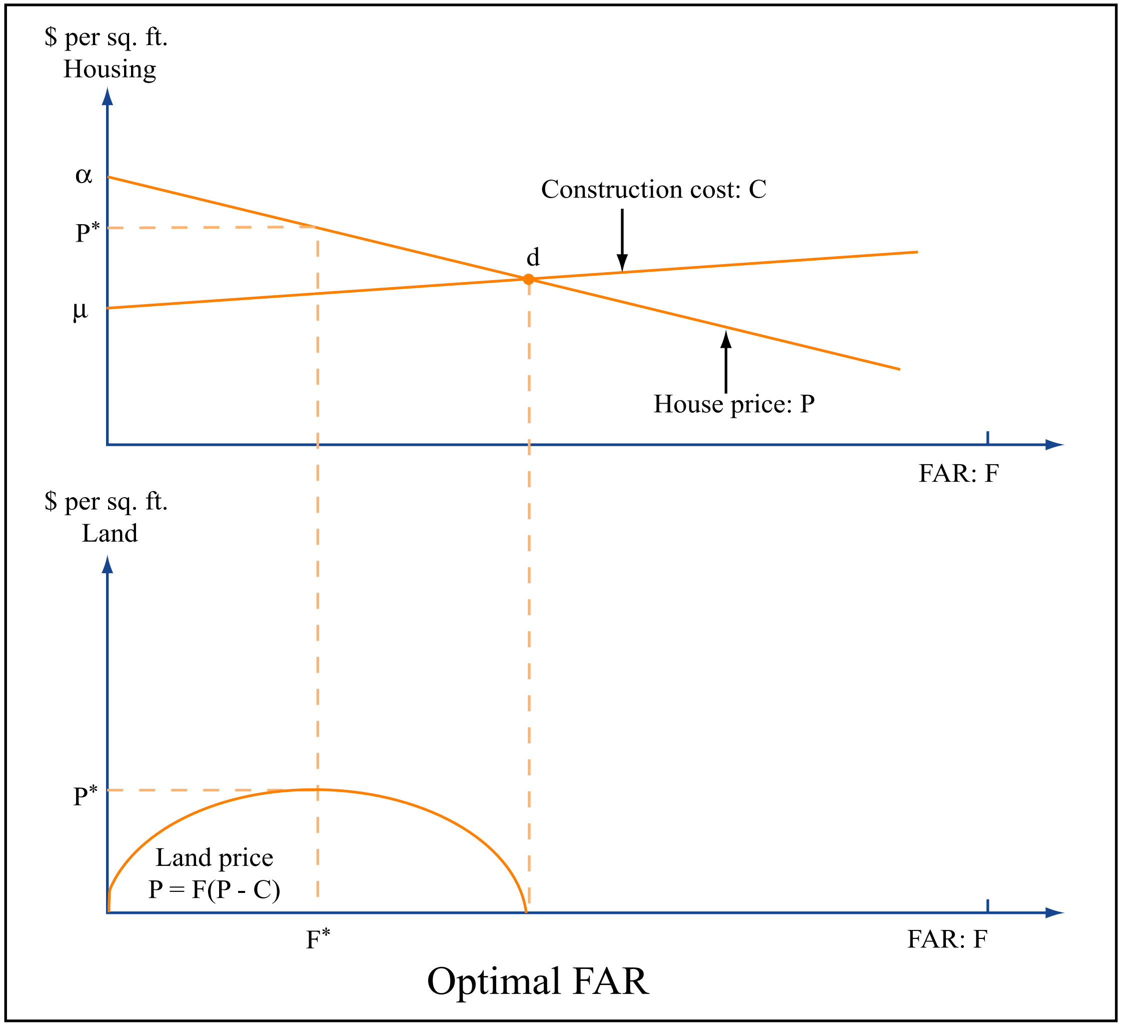


Figure by MIT OpenCourseWare.

What if location value increases?

Optimal density model

House price: $P = \alpha - \beta F$

House construction: $C = \mu + \tau F$

Profit: $p = F [P – C] = F[\alpha - \mu] – F^2[\beta + \tau]$

$\partial p/\partial F = [\alpha - \mu] – 2F[\beta + \tau] = 0$

Optimal density: $F^* = [\alpha - \mu] / 2[\beta + \tau]$,

Maximum land profit: $p^* = [\alpha - \mu]^2 / 4[\beta + \tau]$

Alternative way: think on the margin

Profit is maximized when marginal revenue equals marginal cost

MR=MC

Optimal density model (to be continued in recitation 4)

- Comparative statistics: impact of α, β, μ, τ on F^*
- Location and density
- Factor substitution: land and capital

Refresh your memory: derivatives

$$\frac{dy}{dx} = \lim_{\Delta x -> 0} \frac{\Delta y}{\Delta x}$$

Let y = *f*(x) define a function of *f*.
If the limit exists and is finite, we call this limit the derivative of y with respect to x.

Some simple rules:

- The derivative of a constant is 0
- If $y = x^n$, then $dy/dx = n * x^{n-1}$
- d(cu)/dx = c* du/dx
- At the maximum/minimum points, dy/dx = 0 (the tangent line)