Hedonic Methods and Housing Markets

Chapter1: An Introduction and Discussion of Origins

Edward Coulson
Department of Economics
Penn State University
fyj@psu.edu

*One of the most important tasks now confronting land economics is a reformulation of fundamental theory with respect to land in the light of statistical data. A great mass of accounting and statistical data have become available, and should be employed to determine the facts.*

--L. C. Gray (1928)

## 1.1 Introduction

Real estate markets are important. Shelter is among the most basic of commodities, and one of the things to which much of our income is devoted. Real estate comprises the bulk of a typical US households asset portfolio, and real estate cycles are quite influential on the aggregate US business cycle. Local real estate prices are among the surest barometers of the vigor of the local economy. We discuss few economic issues as readily as the market for real estate in our city, with the possible exceptions of the stock market and gasoline prices. But gasoline and IBM common stock are comparatively homogenous commodities, which makes them comparatively easy to discuss. Every gallon of 89 octane gasoline is basically like every other, and every share of IBM stock is very much like the other. And so when we read in the newspaper that IBM common stock is selling for $50 per share, we can be pretty certain that, for the time being anyway, that our share of IBM stock is worth the same. Not so with real estate. When we read in the newspaper that home prices where we live are going up by 5% per year, it's informative, but such information is limited in the extent to which it applies to particular properties, like one's own; when we find out that the house down the street sold last week for $200,000, that also is informative, but in a different way, and again it's only limited information about the value of our own property.

The information is limited because *every piece of property is different.* This *heterogeneity*

that is endemic to the commodity we call real estate makes the creation and transmission of price information problematic. Knowledge about what else is happening in the real estate market is valuable only to the extent that (say) one's own house has something in common with the houses that created that information.

This monograph is about real estate prices. More particularly it is about how to distill the informational content of individual housing prices into something that one can use to analyze real estate markets and make inferences about them. By necessity this means taking the heterogeneity of housing markets and turning it into something homogeneous. There are two basic methods to do this.

## 1.2 Hedonics

Hedonic analysis is the study of the relationship between the price of a product and the characteristics of that product. People buy and consume residential real estate, through ownership or rental arrangements, because they obtain utility– that is, enjoyment, from the things that the housing unit has to offer, what will sometimes be generically referred to as "housing services". Every house or apartment or duplex has certain attributes, or characteristics (the terms can be used interchangeably), that allow people to obtain utility from residing in it. These include the land, the amount of interior square feet, the age of the house, the number, size, and type of the various rooms of the house, and the existence (or lack of) other amenities such as a garage, air conditioning, source(s) of heat, and the like. Furthermore, because housing itself is immobile, the location is important too; more particularly, the attributes of that location will be experienced by residents as well. The distance of the housing

unit from jobs, shopping and entertainment facilities are examples of location characteristics, and the environmental and sociodemographic characteristics of the neighborhood are all properly considered as attributes of any given housing unit. Each of these attributes contributes to, or detracts from, the ability of the unit's occupant to extract housing services from the unit, and therefore affects the price that people are willing to pay in order to occupy it. The job of hedonic analysis in real estate, then, is to investigate the relationship between the existence and amount of all of these characteristics–structural and locational— and the price that people are willing to pay for the unit. These things can be observed; we can observe the number of bathrooms or square feet a house has or how far it is from the local elementary school, and we can observe the price of the housing unit itself, although the way in which the price is observed can vary (see below)[1]. What is not observable, at least directly, is the way the former are combined in the housing market to create the latter.

It is important to emphasize this obvious point: the housing market, through the actions of the participants, does the combining, through the mechanism of supply and demand. Every housing unit has a distinct price, determined in part by the overall supply and demand conditions in the local housing market, but also by the distinct collection of attributes it embodies. It is the job of hedonic analysis to find the combination method– to find the *hedonic function* which uses attributes as inputs and forms the market price of the unit as its output. Hedonic analysis does this by supposing that each attribute has a market of its own, which is in

[1]Throughout this monograph, the "price" is typically the sale price of the housing unit (sometimes called the *value).*. However it can also be the rent (monthly or otherwise), with little change in the logic or interpretation of the hedonic analysis.

turned governed by supply and demand of its own. Each characteristic therefore has a *hedonic price*, although "price" has a slightly different meaning in this context, as we shall see.

There are at least three reasons why it is important to gain knowledge of the hedonic function in a housing market, or to know the hedonic price of an attribute: (1) One can, for purposes of appraisal and assessment, *predict* the sale price of a housing unit before it is sold; this is possible to do because one can observe the attributes and then estimate the contribution of each to the overall price of the housing unit; (2) for purposes of cost-of-living comparisons and the like, one can create *constant-quality* housing price indexes across time and location. It can be difficult to ascertain the real difference between housing costs in (say) Los Angeles and St. Louis because not only are the overall supply and demand conditions different between the two cities, but the mix of housing attributes is different as well. Houses in Los Angeles might be larger but the plots of land on which they sit might be smaller. In order to know the true cost differences, one must separate out the quality differences. Knowledge of the hedonic price function allows one to do that, and create housing price indexes that account only for the difference in overall market conditions in Los Angeles and St. Louis. (3)For purposes of policy analysis, one can compute the value of particular attributes; thus if one wishes to know the social costs of (say) air quality, hedonic analysis allows the estimation of the "price" of air quality through evaluating the effect of air quality on housing prices. An extension of this brand of hedonic analysis uses such knowledge to compute the net benefits, or consumer surplus, of particular attributes.

We turn, then, by way of introduction, to a brief discussion of these uses, intertwined

with some historical background.

*Appraisal and assessment*

Real estate, i.e. land, and the structures which sit upon the land, is natural fodder for the tax collector. It doesn't move (at least not very quickly) and tax evasion is difficult because it is reasonably easy to find out who owns it. The problem with using real estate as a tax base is that for the sake of fairness the basis of taxation ought to be the market value of the property in question. However, the value of any asset is to some extent unknown unless it is bought and sold. For most assets this is not a problem. When exactly comparable assets are bought and sold, the value of untraded assets can be ascertained; shares of widely-held companies can be valued at the market price on the stock exchange. The price of gold bars is readily obtained. Real estate, however, is different. Because every piece of property is different, one cannot immediately ascertain the value of untraded assets through knowledge of recent trades that have taken place. And no piece of property trades hands with enough frequency to be able to know its idiosyncratic value, or the change in that value from year to year.

In less sophisticated times, property appraisal would, as a consequence, be based on some single characteristic of the parcel or structure, in order to avoid the issues raised above. However there are numerous stories of how reliance on a single characteristic for the purposes of wealth taxation skewed the assembly and/or production of real estate. Plantations along the Louisiana stretch of the Mississippi River, and houses along the canals of Amsterdam were taxed according to the single characteristic of water frontage, and the result was really skinny parcels of land. In England and Ireland, and other countries of Europe, the introduction of the

luxurious commodity of glass into residences led to a "Window Tax" as a way of taxing wealthy homeowners. Windows were a convenient tax base-- they were observable without entering the house, and they were to a certain extent correlated with the ability to pay taxes. But windows are not inelastically supplied, and the window tax led to unusual distortions, such as buildings constructed without any windows, and existing windows being boarded up. While the window tax existed during the Age of Enlightenment, neither the tax, nor the houses, were particularly enlightened.

Therefore it came to be recognized that in order to make property tax collections "fair" and "efficient" one needs to *appraise* the entire value of the property-- that is, define its value in the absence of a market transaction. Again, the idiosyncracies of individual unit make this a problem of a different order altogether than the appraisal of assets like common stock or gold bars. Thus real estate appraisal is nearly as old as real estate taxation, which is in turn at least as old as real estate itself.

Financial institutions also need accurate appraisal of homes on which they make loans. When a prospective borrower needs a loan to purchase land or housing, the collateral for the loan is typically the unit itself. While the secondary mortgage market, (and loan-to-value ratios that are less than one) limit the damage to lenders from default, it is still obviously desirable that the value of the collateral be known (and somewhere close to the actual transaction price). Therefore an evaluation of the property, independent of the potential sale price, is worthwhile and valuable.

A "scientific" method of property valuation was seen as being preferred to one which

had no clear relation to property value itself, the latter seen as the appropriate basis for taxation[2]. Discussion of this set of issues made its appearance in the mainstream economics literature at least as early as the work of Lutz in 1910 (who called these problems "old"). Lutz describes the "Somers System of Real Estate Valuation" which had a number of features worth pointing out: (1) that "real estate values are community values. Land is worth whatever the community thinks it is worth." (2) The mapping of characteristics into value is multivariate. While the street frontage is the characteristic of first importance, other characteristics, including the depth of the parcel, the finish of the building, its age, condition, dimensions and other architectural characteristics are to be included in the valuation; (3) the relationship between characteristics and value is nonlinear. The specific example of this is that the "curve of value" that relates the value of the property to the depth of the parcel increases at a decreasing rate. To the extent that this example represents standard early appraisal methods, appraisers have long known about a number of issues that concern us even today.

*The prices of characteristics*

The real origins of hedonic analysis occur when researchers proposed methods that systematically used data on existing products to derive a statistical relationship between real estate prices and real estate characteristics. One assumes a relationship of the following sort:

$$P = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_k X_k$$

[2]Lutz (cited below) discusses the "gross inequalities [which] have entered into the assessment of real estate"

where $X_1$ through $X_k$ represent the k attributes of a piece of real property (internal square feet, external square feet, number of bathrooms, etc.) and the $a_j$'s are variously referred to as the weights, or coefficients for each of the X's, and P is the price of the property[3]. This is a *hedonic function*. Equation (1) assumes that the hedonic function is linear, but we will later entertain the possibility of nonlinear relationships. Hedonic analysis thus involves using statistical procedures (of which much more in Chapter 2) to calculate values of the $a_j$'s which obviously give the influence of each X on the price.

The first implementations of such techniques involved the pricing of agricultural land. Colwell and Dilmore (1999) have unearthed what is apparently the earliest use of statistical techniques to estimate the contribution of product characteristics to product price. G.A. Haas (1922) used *regression analysis* (about which much more in Chapter 2) to generate a model of farmland prices in Blue Earth County, Minnesota. It is worth discussing Haas's analysis (via Colwell and Dinmore (1999)) despite its early vintage, because it is an interesting introduction to what hedonic analysis does.

Haas theorized that (given his data) the price of land could be appraised by the equation

$$P = \alpha_0 + \alpha_1 B + \alpha_2 L + \alpha_3 S + \alpha_4 U$$

[3]In representing the price of housing, we leave aside whether the price in question is the rent (monthly or otherwise) or the actual sale price of the housing unit. Most of the time the logic and interpretation of hedonic models does not depend on this distinction although the size of the $a_i$'s will depend on which price is being used.

where

*P*= the price of an acre of land (normalized by Haas)

*B*= the (depreciated) cost of the buildings on the land

L= land productivity index

S= Soil productivity index

U= distance in miles to the city center

Thus the four variables B, L, S and U, (playing the roles of $X_1$ through $X_4$ in equation 1) are theorised to be the four things driving the land market in Blue County. Using a set of actual transactions from the years 1916 through 1919, he employed regression analysis to estimate the four weights $\alpha_1$ through $\alpha_4$ as well as the intercept term $\alpha_0$. What he got was this:

$$P = 57.785 + 1.067\,B + .7279\,L + .1685\,S - 3.422\,U \qquad (1.1)$$

Thus, for example, every mile further distant from the city center lowers the price of one acre by $3.42. Thus in a very real sense the "price" of distance is $3.42/mile. Similar things can be said about the other characteristics; moreover, any given piece of land can be appraised for taxation or financing purposes by using the equation above and inserting the appropriate values for the characteristics[4].

[4]Given the computing machinery available to him, four property characteristics would seem to be the limit of the complexity of the model. Thus Haas "adjusted" the property prices in his data set according to the size of the nearest town, and other characteristics. According to

Since this analysis was published only as a technical report of the University of Minnesota (and as Haas's master's thesis) its influence is perhaps fairly small. A similar, yet better known analysis of this type, this time for Iowa, is by Wallace (1926)[5]. Like Haas, Wallace used regression analysis, and used characteristics such as the productivity of the land, and its crop allocation, as well as, in Haas's case, the distance to the nearest town. The fact that he also obtained a negative coefficient makes he and Haas the first in a long line of researchers to empirically validate the basic transport cost model of land prices.

Colwell and Dilmore speculate on the extent to which these papers might have influenced Andrew Court[6]. Court (1939) is generally regarded as the founding father of hedonic analysis, in part because he was the first to use the term "hedonic" in describing his methods. Court investigated the influence of quantities of characteristics such as horsepower and wheelbase on the prices of cars. As Goodman (1998) notes, Court's work would also be recognizable as hedonic analysis today. What Court wanted to do was to break down the price of an automobile into price components, where each component is the measurement of some characteristic of the car. Each component gets a weight, and the weighted sum is the (putative) price of the car, similar to the issue that we have been describing for housing. That is, he

---

Colwell and Dinmore (1999) his adjustments were remarkably close to what he would have obtained by using such characteristics in the regression itself.

[5]Wallace notes that "[e]xperienced land appraisers seem for the most part to disparage scientific procedures". The same thing sometimes happens in this century, too.

[6]Sheppard (1999) also cites the work of Waugh (1929) as another early use of multivariate regression analysis to assess the contribution of characteristics to price, in this case vegetables.

wanted to create a hedonic function for automobiles of the form of equation (1) in which the price of the car is seen therefore as a weighted sum of the value of the various components of a car– a linear hedonic function. Court's ultimate goal was to suggest prices for new automobile models and a new car could, using the above, be priced by merely plugging the component values for the model. Thus could cars be "appraised" before having actually been bought or sold.

Despite the natural application of such methods to residential or commercial property valuation, appraisers did not latch on hedonic analysis as an aid to their task. Neither did economists. As Goodman (1998) notes, hedonic analysis lay dormant in the two decades following the publication of Court's work. It was not until the late 1950s that Zvi Griliches (1958, 1961; see also Fischer, Griliches and Kaysen (1960)) revived not only the methods of Court, but his use of the term 'hedonic'. Griliches analyzed quality change and the construction of price indexes for cars, as well as for other products (even fertilizer).

However, the use of hedonic regression analysis in housing markets only begins with the pioneering papers by Ridker and Henning (1967) and Nourse (1967). Although these authors still do not use the word "hedonic" they are clearly familiar with and are influenced by the work of Griliches.

Ridker and Henning (1967) used averages of housing characteristics and prices by census tracts in a regression framework like (1). However, this study marks a shift in the emphasis of hedonic analysis from appraisal of new or otherwise unpriced products (as was the case in the above-cited papers) to the interpretation of individual hedonic weights. Their

primary purpose was to determine the hedonic price of air pollution levels in the St. Louis metropolitan housing market; they find that increases in the air pollution level of a census tract do have a detrimental impact on the median property values of that tract. A specified decrease of airborne sulfation is found to increase property values by about $245 for each house. That is, the hedonic price of one unit of air pollution is -$245. Then, in the pioneering effort to use hedonic models to analyze social policy, they multiply that figure by the number of households in the entire St. Louis area and argue that the aggregate benefit of the reduction in air pollution would be over $80 million. They further speculate that the cost of such a reduction would be in the neighborhood of $10 to 15 million, so that it would on that account be a net social benefit to undertake the reduction.

Oates (1969) similarly used averages-- this time within communities-- of housing and public expenditure characteristics, again to determine the value of a pair of hedonic attributes-property taxes and education expenditure, in order to test the extent to which these are folded into property values (see Chapter 3).

All three of these studies had the unfortunate aspect that they used averages across areas as the units of observation. For obvious reasons the use of data on individual houses is preferable, although quite difficult to get at that time. A sense of the difficulties faced by researchers can be gained through reading what is apparently the first hedonic analysis using data on individual houses, the paper by Kain and Quigley (1970). They use data from St. Louis on about 1,500 houses and their neighborhoods which was gathered through door-to-door surveys. They use the word hedonic, identify the coefficients as prices of the characteristics,

and note the usefulness of studies like theirs for the appraisal process and for the construction of constant-quality price indices. These authors provide a sophisticated analysis of the relationship between different housing attributes and use factor analysis to sensibly respecify the regression model.

From these beginnings, including also the early papers by Anderson and Crocker (1972) and Grether and Mieskowski (1974) among others, hedonic analysis became a standard tool in the analysis of housing and many other differentiated commodities.

In Chapter 2 of this monograph we cover the construction and evaluation of hedonic price models, and in Chapter 3 we discuss the particular problems and opportunities associated with particular classes of attributes.

*Price indexes*

The second common use of hedonic prices is in the creation of price indexes. A price index is a measure of the "average" house price in a given location at a given time. The comparison of "average" housing prices in different cities, or in the same city at different time periods (or indeed for different cities in different time periods) is fraught with danger, again because the times and places one is comparing have different housing stocks. Hedonic analysis can help alleviate this problem because it provides a means for comparing two houses with identical characteristics. Using the methods of Chapter 2 , one can create hedonic models for the different locations and/or time periods and then insert an identical set of (average) X characteristics into the model, and you will have an estimate of the difference in prices in the two markets. In Chapter 4 we explore this procedure in more detail.

*Repeat Sales*

With the advent of huge databases of property assessments and/or transactions, it has become possible to implement a different methodology for estimating indexes. *Repeat-sales* simply look at pieces of property that have more than one sale. By looking at the rate of price increase for those houses for whom a price increase can be calculated (because it had more than one transaction) and averaging these over all such houses (taking into account the different points in time in which these transactions occurred) one obtain a (temporal) price index (or something very much like it) without having to take into account the characteristics of the housing stock. Such repeat sales models will also be covered in Chapter 4.

*Technical Material*

Chapter 5 explores the theoretical underpinnings of the hedonic price model. These underpinnings are complex, and the chapter will only introduce the subject. We then take up the interesting question of how to estimate the demand for hedonic attributes. This is also, of necessity, relatively technical, but is of substantial interest for the evaluation of public policy. Standard economic reasoning tells us that only through estimating the demand curve can we calculate the true benefits (i.e. consumer surplus) of environmental goods such as air pollution. Chapter 5 closes with some newer modeling techniques designed to estimate such surpluses.