Two-Tier Frontier and Generalized Kernel Estimation of Hedonic Price Indices

BY

Christopher F. Parmeter

BA, Nazareth College, Rochester, New York, 2001

MA, Binghamton University, Binghamton, New York, 2003

Dissertation

Submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy in Economics

in the Graduate School of

Binghamton University

State University of New York

2006

UMI Number: 3213277

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Accepted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy in Economics

in the Graduate School of

Binghamton University

State University of New York

2006

May 9, 2006

Daniel J. Henderson, Department of Economics, Binghamton University

Subal C. Kumbhakar, Department of Economics, Binghamton University

Solomon W. Polachek, Department of Economics, Binghamton University

Abstract

This dissertation investigates the development of hedonic price indices both theoretically and empirically. The recent advances of nonparametric statistical methods lend themselves to hedonic price estimation where a mix of continuous and discrete attributes exists. Formal considerations of recent insights as to the underlying specification of hedonic prices are also considered. Extensions of hedonic price models to incorporate search and bargaining are considered. Aside from the discussion of the two stage hedonic estimation method in chapter 1, the remaining chapters deal with first stage estimation issues only.

To the most important women in my life, Sandra and my Mother.

Acknowledgments

I would like to thank my advisors Dan Henderson and Subal Kumbhakar for putting up with me during my stay here at Binghamton. They have shaped my thoughts and influenced my opinion of the way economic research should be conducted to a degree that I cannot explain. I owe them a dept of gratitude that can only be paid back with my thanks and hard work in the years to come. I would also like to thank Solomon Polachek. Without his influential insights and creative models my research would have fallen short in preparing this dissertation. Thanks for Anton Schick are in order as well. It is with great appreciation that I acknowledge his kindness for 'allowing' me to attend his statistics classes. I am sure that my expression intimated the difficulty of the material, and yet he was always willing to help me with any questions that I had. I would also like to thank Barry Jones and Clifford Kern for many helpful discussion along the way and to Thomas Cowing for patiently recruiting me to Binghamton. I have enjoyed everyday I have spent in the Southern Tier and am sad to go.

I would also like to acknowledge the support I have received from my family and friends. They have stuck by my side even thought there were countless weeks and months were they did not see me. I appreciate their understanding for my desire to obtain my Ph.D. and hope that they know how much I truly care for them. Special thanks are due to Jared Strief and Jeffrey McCormack. I know that it must have been hard for them to go from seeing me all the time in college to receiving an occasional phone call from me once or twice a month. I cannot thank the two of them enough for their unwavering support. Michael O'Hara deserves just as much credit for the completion of this dissertation as anyone else. He was always willing to stop his work to listen to me ramble about my research and he has been a first class office mate and fried for the last three years.

Table of Contents

List of Figures

List of Tables

Chapter 1

Survey of Hedonic Price Index Theory and Estimation

1.1 Introduction

The evolution of interpretation and estimation of hedonic price indices has kept the attention of economists, environmentalists, urban planners, and econometricians alike for over 40 years. The state of the art has changed dramatically over that time and the landscape of the theory has evolved considerably. It is the aim of this chapter to summarize the changes and contributions to the hedonic price literature that have taken shape over the course of more than four decades of intense economic research. Given recent advances in the estimation and understanding of hedonic models, this chapter presents historic as well as current ideas on the interpretation and estimation of hedonic price indices. Both theoretical and econometric advances will be summarized as well as an overview of where the literature has gone over the last four decades since the explosion of economic interest in the subject.

Imagine you were given the task of determining the price of and demand for a certain good that is not traded in a market. Lets call this good the non price good. Quite a task at first glance. Now suppose that you had access to another good whose price is observed and is composed of the non price good. Lets call this good the observed price good. With enough data on the observed price good, you can determine the price of and demand for the non price good using a hedonic price index. In fact, many an urban planner and environmental economist use hedonic price indices to estimate the price and demand for goods such as proximity to an opera house, air pollution, noise from a highway, or the view of the mountains.

Taken literally 'hedonism' is defined as "living and behaving in ways that mean you get as much pleasure out of life as possible, according to the belief that the most important thing in life is to enjoy yourself"[1]. The principle of utility or profit maximization discusses the behavior of agents as acting in their own best interest, to do whatever pleases them. Given this economic agents can be viewed as hedonists. That economists use and study hedonic prices should come as no surprise then, as economic agents are willing to pay for goods that make them happy and provide pleasure. Thus it would seem that the study of hedonic prices is at the core of economic theory.

The study of hedonic prices was originally intended for the construction of price indices that placed a "market price" on goods where no market existed. [2] In the hedonic index framework, price was regressed on the utility bearing attributes of a specified product to determine each attribute's implicit price, i.e. the price per unit of that specific attribute the consumer (producer) was willing to pay (accept) to acquire (sell) more of it. Thus, hedonic price indices were constructed so that price changes could be decomposed into those relating to qualitative changes in the good in question (were cars more expensive because they were of higher quality?) and to changes that were indicative of the economy (were cars more expensive due to inflation over a specified period of time?).

What the hedonic method provided was a means to compare different products in different markets. For example, apartments in New York City may rent for $1,000 a month compared to apartments that rent for $500 a month in Upstate New York. Does this imply that those living in Upstate New York are getting a good deal and those in New York City are overpaying for their living arrangements? The answer lies in the attributes of the apartments that each is living in. If both apartments were identical then, yes, the New York City tenant would be overpaying, however, if their attributes were quite different, then it may be that the price differential is completely accounted for. This is the allure of hedonic methods and why economists spend so much time using them and investigating their properties.

[1] Cambridge Dictionary

[2] The first hedonic study (see Court (1939)) was in the automobile market where the non market goods were the attributes of automobiles, things such as horsepower, weight, and wheelbase. Waugh (1929) looked at quality in the vegetable market, however, the intuitive aspect of vegetables being differentiated products with varying attributes is much harder to grasp than that of automobiles.

The remainder of this review is as follows. Section 1.2 briefly reviews the theory and estimation of hedonic price indices prior to Rosen's seminal piece in 1974. Section 1.3 discusses in detail the economic model developed by Rosen that captures both sides of the market in a hedonic price index as well as discussing Rosen's proposed two-stage method that can uncover structural parameters and Section 1.4 reviews the econometric implementation of Rosen's method that took place in the late 70's and early 80's. Section 1.5 discusses the econometric critiques of Rosen's two stage method that occurred throughout the 80's, while Section 1.6 discusses some interesting insights regarding the estimation and understanding of hedonic price models that have just recently come to light. Section 1.7 contains a conclusions and discusses further extensions of the Rosen model and areas of hedonic theory that need more investigation.

1.2 The Evolution of Hedonic Price Indices Up to 1974

The underlying theory for hedonic models was that of differentiated products. Goods that by name seemed identical, such as "house" or "car", were actually composed of attributes that significantly differed from good to good and it was desired to determine how their prices were effected by this differentiation.[3] While Court (1941a,b) was the first to rigorously develop a mathematical model to study differentiated products across an infinite spectrum of attributes, it was not until the contribution from Houthakker (1952) that the field's understanding of consumer behavior and the link to quality variation was brought to the fore. Improvements and enhancements to Houthakker's model were added by Becker (1965), Lancaster (1966) and Muth (1966).

Tinbergen (1956) was the first to formally apply this theory, looking at the labor market. The idea that a worker could not be unbundled and spread across many tasks simultaneously led to a hedonic wage regression, where workers' were differentiated by their skill sets rather

[3] In our setup the attributes are the non price goods and the product in question is the observed price good.

than being viewed as a homogeneous good that should command the same price in the market. Tinbergen discussed the interaction between the distributions of worker and firm characteristics as well as the parameters of the utility and cost functions in the determination of a hedonic wage function. Tinbergen was also able to derive an analytic equilibrium hedonic price function assuming all producer and consumer parameters were distributed normally and that both utility and cost functions were quadratic. If one further assumed that the slopes of the bid and offer functions were linear then this model becomes the now famous Normal-Quadratic-Linear model (or NQL for short). As it would turn out to be the case later, this model became one of the most widely used and heavily criticized (by econometricians and empiricists alike) versions of the hedonic price model.

Another of the brilliant insights of the hedonic price method is that it turned into a valuable non-market valuation technique. Because the attributes that were implicitly being priced could not be taken out of the actual good that was sold, no formal market existed for them. You would not take your car to the horsepower market and buy more horsepower to put into your car, you would have to buy a different car to get more horsepower. So, these indices were truly unique in that they represented a way to attach value to goods that were not common economic goods that were bought, sold, and traded in typical markets. This insight was picked up by environmental and urban economists in the 1960's and 1970's as a way to value improvements in neighborhoods, structural improvements to buildings, as well as environmental amenities such as noise, distance to parks, and air pollution levels. See de Leeuw (1971) and Ball (1973) for reviews of hedonic housing studies and Freeman (1979) for a survey of air quality valuations using hedonic price theory.

Griliches (1961, 1971), Ridker and Henning (1967), Kain and Quigley (1970), and Gordon (1973) were the first, well known papers, to empirically employ hedonic price functions, primarily for the construction of price indices. These indices were constructed from the attributes of the good for which the price index was being created. The idea behind the hedonic evaluation of goods was that no formal market existed for the attributes of the good and so direct knowledge of their value was unknown. Collecting data on a goods price, as well as the attributes it was composed of, allowed empiricists to uncover the value that any particular attribute had on the overall price of the good.

However, once this hedonic price function was estimated, several studies tried to determine the aggregate benefits that would accrue to buyers (sellers) given a change in one or more of the attributes. For instance, Ridker and Henning (1967) tired to determine the effects of an improvement in air quality through its corresponding impact on house values. Unfortunately, their estimates would only be correct if the hedonic price function they estimated was exactly equal to a demand function, as gross benefits are calculated as the area beneath a demand curve. For this to be true required that consumers were **identical**, something not likely to hold. Several comments were written after the Ridker and Henning study that criticized the method on several grounds.

Freeman (1971) noted that the coefficient from the hedonic regression was composed of both demand and supply factors and left little room for interpretation. He states on page 415, "... this relationship is the result of the interaction between the availability of land with different levels of air quality (supply factor) and tastes and preferences, other prices, income, and its distribution (demand factors). For any given set of demand factors different supply factors will lead to different patterns of property values and different regression results." This was possibly the first recognition that the hedonic price function represented an equilibrium between buyers and sellers and that coefficient interpretations required care and considerable thought. Freeman went on to suggest that a formal theory that took this interaction into account be developed before researchers attempted to use hedonic price regressions to construct aggregate benefits.

Other comments on the Ridker and Henning study followed after Freeman (1971). Anderson and Crocker (1972) pointed out that the coefficient estimates from the hedonic regression represented the marginal bid of a consumer and so could be used to conduct benefit analysis, contrary to Freeman. Freeman (1974a) responded to this comment by noting that the assumptions needed to justify the interpretation of the marginal benefits were quite restrictive and unlikely to hold in real settings. Polinsky and Shavell (1975) and Small (1975) defended Freeman's stance and discussed alternate viewpoints of the model in order to better the understanding of the situation at hand. One point that was never formally raised in any of these comments was the linear form of the air quality measure within the multiple regression framework of Ridker and Henning. This lead to a constant marginal valuation and placed a stringent assumption about the underlying economics that generated the model.

Although theoretical considerations regarding differentiated products existed long before Sherwin Rosen's seminal piece in 1974[4], the simultaneity between consumers and producers was original. Indeed, as Rosen pointed out, the interface between buyers and sellers left one with very little structural interpretation to hedonic price coefficients since they were an amalgam of both sides of the market. This in turn diminished the use of hedonic coefficients being used to conduct aggregate benefits analysis. Instead of viewing the hedonic regression as an index number generator, he instead saw it as a reflection of equilibrium in the market for differentiated products, reinforcing Freeman's (1971) urging for a theoretical model that represents a general equilibrium and allows testable hypotheses. This revolutionized the use and understanding of the hedonic price regression. Indeed, this insight has received over 1200 citations since its publication (Social Science Index, checked March 2006). Rosen's theoretical model and its econometric implementation are introduce next.

1.3 Rosen's Model

The economic model put forth by Rosen (1974) amounts to the description of a competitive equilibrium where buyer and seller interaction takes place over a commodity spectrum. Here the commodities can be thought of as product attributes and there is one good being traded

[4]Contemporary to Rosen, Freeman (1974b) and Mas-Colell (1975) developed models of market equilibrium with differentiated products, but rarely are cited in the hedonic price literature.

in the market. Thus, any product being sold in the market can be characterized by a vector of attributes $z = \langle z_1, z_2, \ldots, z_q \rangle$, where z_i measures the amount of the ith attribute that the product contains. If the product being sold was a television, the z's could measure diagonal screen length, if the television is HD ready, the weight of the TV, if the TV can be wall mounted, etc. In the hedonic price index literature the z's can be thought of as rough measures of quality.

The jumping off point to develop the model theoretically is the assumptions that surround the hedonic market. Here Rosen assumes that all of the attributes are measured objectively so that all consumers share an equal insight into the composition of the product, this would rule out the use of hedonic methods to value products in a market where lemons exist. He also assumes that consumers are fully informed so that they know of the lowest price that exists in the market for the specific product with the desired attribute vector. In line with Court (1941a,b) he assumes that the commodity vectors come from a spectrum instead of being selected out of a finite number of attribute bundles.[5] Lancaster (1966) analyzed the consumer side of the market without assuming continuity of attribute bundles.

Consumers and producers each act as price takers and make their decisions based off of a market hedonic price function which determines the price the product in question sells for given the specific attribute vector it is composed of. What will turn out to be quite important later, the hedonic price function is nonlinear, suggesting that the product cannot be unbundled and the individual attributes sold in separate markets. This implies that there do not exist arbitrage possibilities for consumers or producers to eliminate price differences across the product.

Rosen's model provided the theoretical underpinnings for the hedonic price model, assuming perfectly competitive markets and full information within the market. Indeed, his model was the first to theoretically show that the hedonic price function was simultaneously an envelope for buyers' bid functions and an envelope for sellers' offer functions. What this

[5]Here bundle refers to the vector of attributes that the consumer is purchasing when the product is being consumed; it is used interchangeably with attribute vector.

result suggested was that even with estimates for a hedonic price function, the resulting coefficients had no structural meaning since they simultaneously represented demand and supply. The simultaneous structure uncovered in his model led to the development of a two-stage approach to uncover the structural parameters of interest, namely the utility and cost technology parameters for buyers and sellers, respectively. The following subsections discuss the formal model of Rosen as well as the two-stage estimation procedure suggested by Rosen when closed form solutions for equilibrium cannot be obtained.

1.3.1 The Consumer Side of the Market

Consider a hedonic market where consumers are paired with one and only one seller. Let $P(z)$ represent the hedonic price function that consumers base their decisions off. A representative consumer has utility function $U(x, z, \zeta)$ where x represents a composite commodity reflecting consumption of all other goods and ζ is a vector of taste parameters that characterize the utility function and have joint distribution $f(\zeta)$. In an econometric setting ζ would be composed of those parameters that are observed, ζ_o, and those parameters that are unobserved (by the econometrician), ζ_{uo}. The composite commodity is assumed to have unit price. Rosen followed the traditional beliefs of the utility function, strictly concave and increasing in all coordinates. The consumer's budget constraint is given as $y = x + P(z)$ where y represents income. If $P(z)$ was linear this would be the standard constrained utility maximization problem that composed general equilibrium models, however, given that $P(z)$ is nonlinear the analysis leads to a quite different picture of market equilibrium. Replacing composite consumption within the utility function as $x = y - P(z)$ the first order conditions are,

$$U_z(y - P(z), z, \zeta) - U_x(y - P(z), z, \zeta) \cdot P_z(z) = 0. \tag{1.1}$$

To ensure that the attribute solution comprises a maximum, the corresponding second order condition is (suppressing dependence on specific variables),

$$U_{zz'} - 2U_{zx} \cdot P_z - U_x \cdot P_{zz'} + P_z' \cdot U_{xx} \cdot P_z \quad \text{is negative definite.} \tag{1.2}$$

Rearranging (1.1) shows the fundamental trait of the hedonic price function: the slope of the hedonic price function (in the ith attribute) represents the marginal rate of substitution between this attribute and the composite commodity, holding all other attributes fixed ($P_z = U_z/U_x$). Only in the special case that the marginal utility of the composite commodity is constant does the slope of the hedonic price function represent a classic compensated demand curve. This fact was largely ignored in the subsequent years following Rosen's paper and many of the papers that employed his methodology econometrically misinterpreted their results, basing their results and interpretations on demand functions rather than marginal rate of substitution functions. While the second stage estimation goes through regardless of whether the marginal price represents a compensated demand curve or a marginal rate of substitution curve, interpreting the estimation results incorrectly can lead to poor policy analysis.

That Rosen derived these first order conditions was nothing special. His major characterization of the problem is that he rearranged it into one that captured the spatial context of the market. By fixing utility and income at pre-specified levels and introducing bid functions into the analysis, Rosen was able to characterize the market in a manner that provided a more intuitive understanding of what the hedonic price function actually represented. The bid function, $\theta(z; u, y)$ holds utility and income fixed. It represents the expenditure a consumer is willing to pay for different attribute vectors for a given utility-income index. Thus it traces out a family of indifference curves relating the attributes with forgone amounts of the composite commodity. Incorporating the bid function into the utility function in the same manner as the hedonic price function, $U(y - \theta, z, \zeta) = u$, results in the following first order conditions,

$$U_z(y - \theta, z, \zeta) - U_x(y - \theta, z, \zeta) \cdot \theta_z(z) = 0, \tag{1.3}$$

which almost looks identical to (1.1). The corresponding second order conditions guaranteeing a maximum are,

$$U_{zz'} - 2U_{zx} \cdot \theta_z - U_x \cdot \theta_{zz'} + \theta_z' \cdot U_{xx} \cdot \theta_z \quad \text{is negative definite.} \tag{1.4}$$

Given that both the bid function and the hedonic price function satisfied the same condition, it became evident that in equilibrium consumer's bid functions were tangent to the market hedonic price function. This in turn suggested that the hedonic price function represented an envelope of consumer's bid functions in equilibrium. For $u_1 > u_2$ it is the case that $\theta(z; y, u_1)$ lies everywhere beneath $\theta(z; y, u_2)$. That this is so follows from the fact that with the same income and attribute vector, to achieve a higher utility level the bid must be lower, so as to have more money left over for other consumption. Thus, the hedonic price function represents an upper envelope and equilibrium is characterized by the hedonic price function being everywhere above the family of bid functions that correspond to this equilibrium. An illustration of equilibrium for one dimension of z is presented in Figure 1.1 for three different consumers.

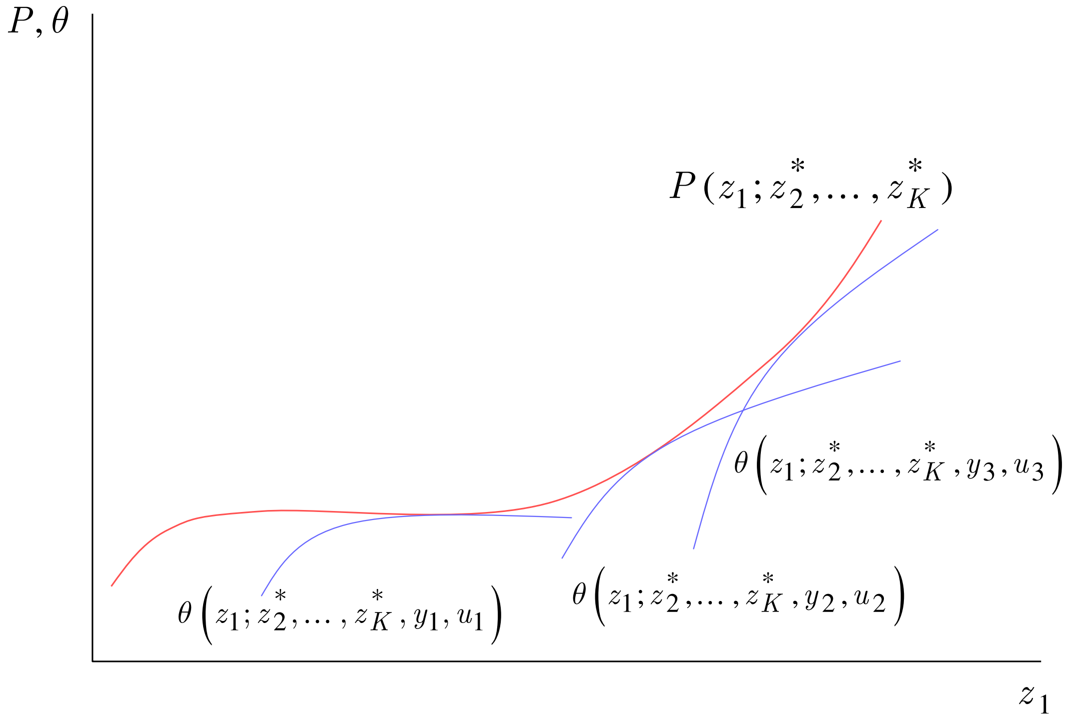

Figure 1.1: Consumer Equilibrium

It is evident from Figure 1.1 that the hedonic price function is more heavily curved than than the bid functions. If this were not the case then the first order conditions for utility maximization would not be satisfied as the tangency points would break down. Another insight that can be gathered from the figure is that consumers with similar tastes and incomes will locate around similar product specifications. This has important implications for hedonic analyses of housing markets in the formation of neighborhoods. Thus, the hedonic price method can explain market segmentation and corresponds nicely with spatial models of equilibrium. Rosen extended the consumer side of the market to include the purchase of multiple units of a product. There consumers had to maximize based on the attribute vector and the number of products to purchase. The analysis of this setup is similar and it does not provide a more intuitive perspective of the problem so we leave the reader to Rosen's paper for more discussion and implication in this framework.

1.3.2 The Producer Side of the Market

Upon analyzing the consumers' side of the market Rosen took on determination of optimality conditions for producers; this was the typical profit maximizing framework. Here firms produced M units of the good in question and took prices as given. Thus, his original model was not suited for analysis of markets characterized by price-setting behavior. Costs for a firm where characterized by the industry cost function, $C(M, z, \upsilon)$, where υ represented cost (production) parameters that varied across producer with joint distribution $f(\upsilon)$. As with consumer taste parameters, υ is composed of parameters that are observed, υ_o and parameters that are unobserved by the econometrician, υ_{uo}. Again, Rosen made typical assumptions regarding the cost function, namely convexity and positive marginal costs for both attributes and number of units. Thus in Rosen's profit maximizing framework with the hedonic price function given, the first order conditions satisfied by producers were,

$$M \cdot P_z(z) = C_z(M, z, \upsilon) \tag{1.5}$$

$$P(z) = C_M(M, z, \upsilon). \tag{1.6}$$

The first order conditions were characterized at the optimum that the slope of the hedonic price function (marginal revenue) equal the marginal cost of production per unit being sold at the optimum and the marginal cost of selling another unit equals the hedonic price at that attribute level (unit revenue). As with the consumer, the negative definiteness of the matrix of second derivatives would ensure a maximum for profits. The corresponding second order conditions are (with dependence on variables suppressed),

$$M \cdot P_{zz'} - C_{zz'} \quad \text{is negative definite} \tag{1.7}$$

$$C_{MM} < 0. \tag{1.8}$$

The symmetry with the consumer side of the analysis becomes apparent with the definition of an offer curve. The offer curve, $\psi(z;\pi)$, is the firm's version of a consumer's bid function. Here firms offer different levels of the attribute vector for a fixed profit. These curves trace out iso-profit relationships between the individual attributes. Here, as with the consumer analysis, the hedonic price function is replaced by firms' offer functions and the subsequent first and second order conditions are derived. Not surprisingly they look almost identical to (1.5) through (1.8).

$$M \cdot \psi_z = C_z(M, z, \eta) \tag{1.9}$$

$$\psi_\pi = 1/M \tag{1.10}$$

$$M \cdot \psi_{zz'} - C_{zz'} \quad \text{is negative definite} \tag{1.11}$$

$$\psi_{\pi\pi} < 0. \tag{1.12}$$

As before, the first order conditions imply that the marginal offer price (at constant profit) is equal to the marginal cost of production while the marginal offer price (at constant attribute levels) is constant, thus, offer functions for different levels of profit with the same attribute vectors have the same slope. For higher levels of profit ($\pi_1 > \pi_2$) a firm's offer function should be higher than for a lower profit level, $\psi(z;\pi_1) > \psi(z;\pi_2)$. Intuitively this means that for the same attribute vector (and cost of production), to obtain a higher level of profit, the firm must offer the good for a higher price. Thus for any given firm, the offer functions lie strictly above one another and the first order condition implies that the hedonic price function represents a lower envelope of these offer functions with equilibrium corresponding to the hedonic price function being everywhere beneath and tangent to the profit-attribute indifference surface.

An example for one dimension of z is provided in Figure 1.2. Here, as opposed to the consumer side, the hedonic price function is less curved than producers' offer functions. In fact, it is this relationship between the curvature of the hedonic price function and the curvature of the bid and offer functions that will allow for identification of structural parameters later down the road. If they were all on top of one another then identification would be impossible. If the curvature were less than firms' offer functions then firms would not be producing at optimal levels and would need to adjust. Even though the results for consumers suggested separation based on specific attributes, here segmentation of producers is not quite as obvious. One interpretation may be that certain types of producers sell certain packages of the goods. That is, those selling cars who are auto mechanics and keep their vehicles in good shape sell a different attribute bundle than those who are not auto mechanics.

The spatial location of consumer and producer has important econometric implications as it suggests that buyer and seller are not paired up at random, but are matched together by the market. This has implications for certain styles or qualities of goods being bought and sold by certain types of buyers and sellers, respectively, creating an endogeneity within the market that is hard to disentangle econometrically. Before discussing the econometrics of the model lets delve into the equilibrium that arises as a result of utility and profit maximization.

1.3.3 Equilibrium/Recovery of Salient Parameters

Equilibrium in a hedonic market is characterized when the solution for consumers is equivalent to the solution for the producers. This would imply that the hedonic price functions given in Figures 1.1 and 1.2 are the same and so at every sale the buyer's bid function is

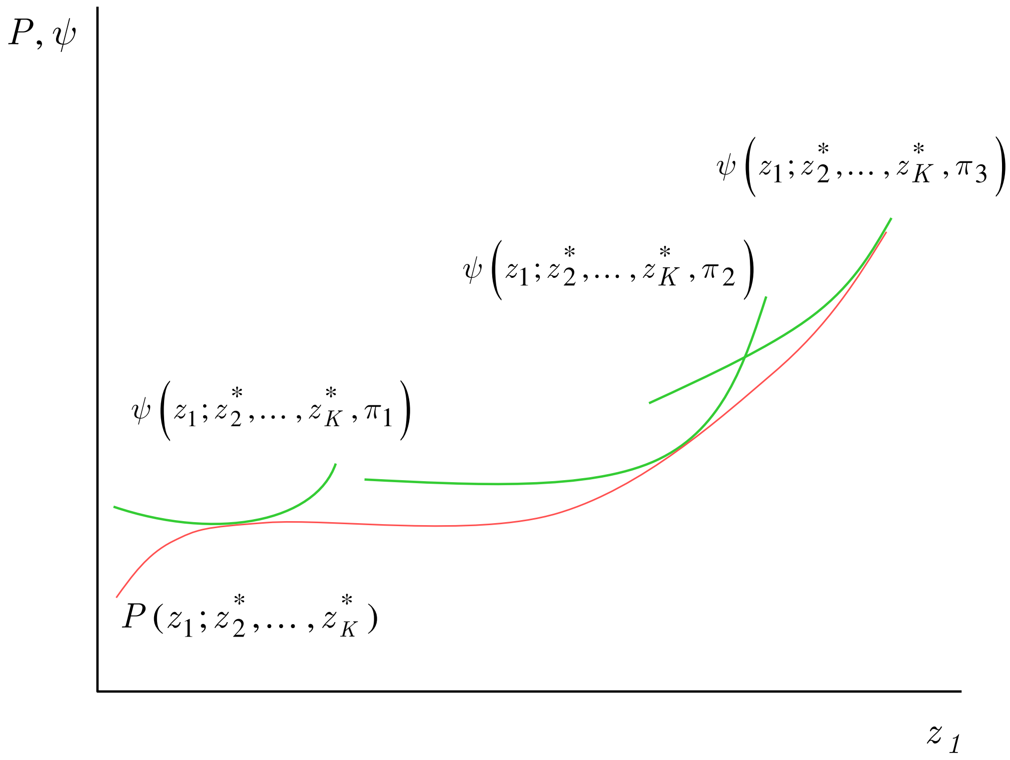

Figure 1.2: Seller Equilibrium

tangent to the seller's offer function, which are both tangent to the hedonic price function. A graphical representation of equilibrium is given in Figure 1.3.

The implications of the curvature of the hedonic price function discussed in the previous subsections become more apparent from Figure 1.3. We see that, in equilibrium, buyers' bid functions have less curvature than the hedonic price function, which in turn has less curvature than sellers' offer functions. In fact, it can be shown that the hedonic price function is a weighted average of buyers' bid functions and sellers' offer functions.

An analytical solution for short run equilibrium was given by Rosen, assuming that there was one attribute which was uniformly distributed across firms over some prespecified range. His specific assumptions led to a second order differential equation that had a known

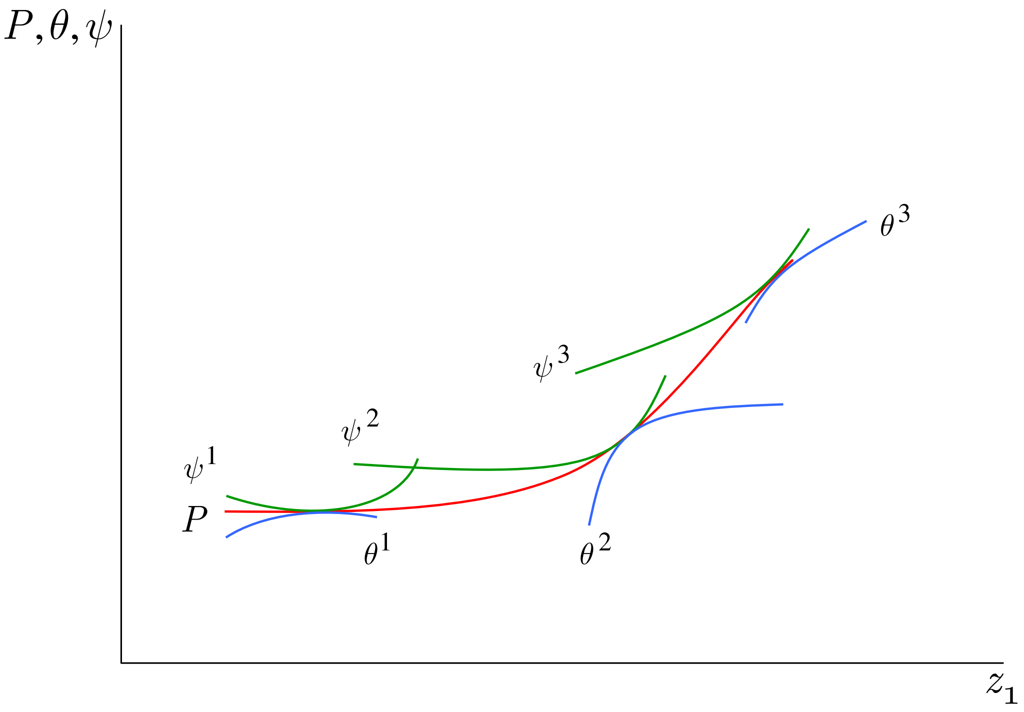

Figure 1.3: Market Equilibrium

solution. However, in either Rosen's or Tinbergen's model, there is no theoretical justification for the use of the specific utility and cost functions, nor for the distributional assumptions of the parameters in those functions. Aside from these two models, closed form solutions to hedonic equilibrium do not exist and so an alternative method was required to investigate these markets.

While the simultaneity of Rosen's model was ingenious and provided great insight into the implications of hedonic price indices, Rosen also showed how the parameters related to consumers and producers could be recovered with explicitly solving the second order differential equation. Knowledge of these parameters would allow researchers to uncover marginal rates of substitution between attributes as well as income elasticities for consumers.

On the producer side marginal rates of technical substitution could be uncovered as well as elasticities of substitution.[6]

Rosen's insight to the recovery of agent parameters was from equations (1.3) and (1.9). The derivatives of consumer's bid functions were proportional to marginal rates of substitution between an attribute and the numeraire good. In equilibrium the this marginal rate of substitution function must intersect the marginal price of the hedonic price function. Unless the marginal utility of the numeraire good is constant these marginal rate of substitution curves are not compensated demand curves, as many would claim later on. Similarly, the slope of producers' offer functions reflect the reservation supply price for the attribute and is proportional to the slope of the hedonic price function (which may be thought of as the marginal revenue function) in equilibrium. This description of equilibrium is presented in Figure 1.4.

Figure 1.4 presents equilibrium in a standard way: supply equals demand. Here however we have marginal rates of substitution instead of demand functions. From Figure 1.4 one can see the shortcomings of the standard hedonic method for uncovering structural parameters. Using the coefficient estimates from a regression of price on attributes would understate the slope of the marginal rate of substitution functions and overstate the slope of the compensated supply curves.[7] So another method would be needed if the structural parameters of interested were to be uncovered in an unbiased fashion.

Rosen used the intuition from Figure 1.4 to suggest an econometric technique to estimate the structural parameters that composed the compensated supply and marginal rate of substitution functions. His procedure went in two steps. Consider the following system of

[6]The cost side recovery was not as important for Rosen's method as was the consumer side. This was due to the fact that standard production econometrics allowed one to estimate these parameters. Since utility is unobserved no formal econometric method existed prior to Rosen's suggestion.

[7]This was noted by various authors who applied Rosen's two-step method which will be discussed later on.

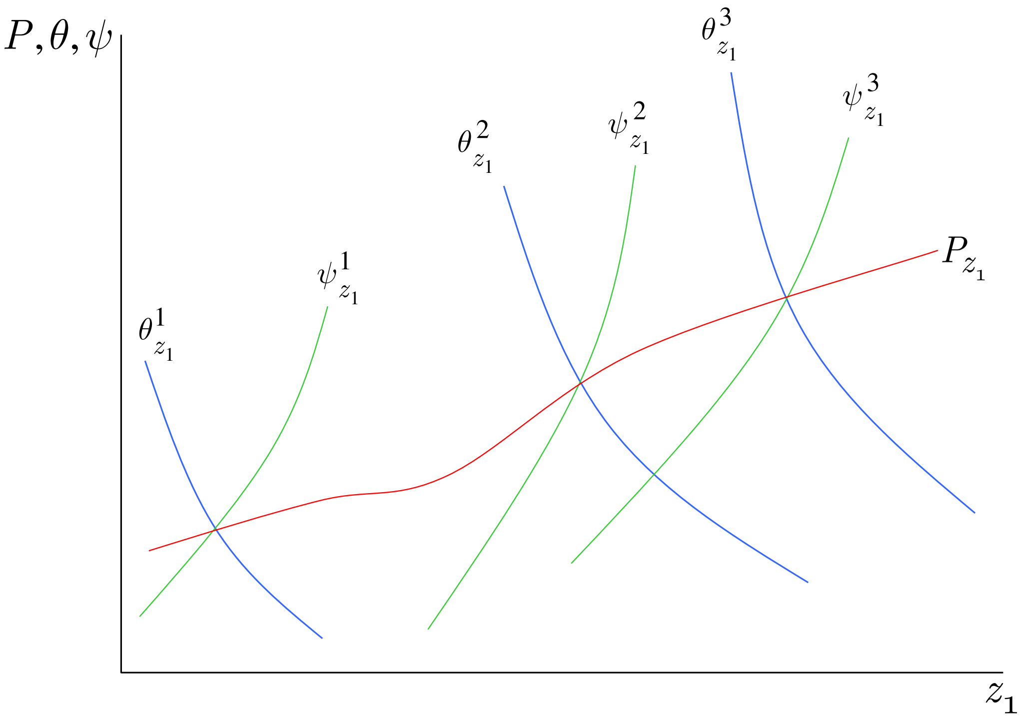

Figure 1.4: Alternative Depiction of Market Equilibrium

equations.

$$P = h(z_1, \ldots, z_q) + \varepsilon^h_{uo} \tag{1.13}$$

$$P_z = \theta_z(z_1, \ldots, z_q, \zeta_o) + \varepsilon^h_{\zeta_{uo}} \tag{1.14}$$

$$P_z = \psi_z(z_1, \ldots, z_q, \upsilon_o) + \varepsilon^h_{\upsilon_{uo}}, \tag{1.15}$$

where P represents the observed price, P_z is the vector of slopes of the estimated hedonic price function, ζ_o is a vector of observable buyer attributes, υ_o is a vector of observable producer characteristics, $h(\cdot)$ is the hedonic price specification chosen by the econometrician, $\theta_z(\cdot)$ is the buyers' marginal rate of substitution function derived from the first order conditions and ψ_z is the sellers' supply functions from their corresponding first order optimality conditions.

Our regression errors, ε^h_{uo}, $\varepsilon^h_{\zeta_{uo}}$, and $\varepsilon^h_{\upsilon_{uo}}$, come from the unobserved taste and technology parameters. Rosen's procedure went as follows:

1. Estimate (1.13) using whatever method you like.
2. Estimate the system of $2q$ equations given in (1.14) and (1.15) replacing P_z with $\hat{P}_z$ obtained from the first step.

The bulk of the criticism levied towards Rosen's two step procedure has been estimation of the second step. However, Rosen never commented on the econometric aspects of this method except to suggest that it represented a 'garden variety' identification problem since there were $2q$ equations in the system and only q unknowns. Much of the research that has commented on the weaknesses or holes in Rosen's method have actually been the results of empiricists attempting to implement his method in an econometrically incorrect manner. Even though Rosen's method has been criticized it still remains the subject of much debate and intrigue, given its established and highly cited reputation within economics.

Given that one of the most important aspects of hedonic price analysis is to uncover the true structure of demand and supply, the subsequent research on hedonic price estimation took two paths, one of empirical implementation of Rosen's method to uncover the structure of the economy, and another of investigating the econometric implications of the two-step method put forth by Rosen to determine the validity of the empirical applications. The following two sections will look into applications of the two step method and the econometric criticisms levied toward the method.

1.4 Early Applications of Rosen's Two Step Methodology

Most of the 70's was spent estimating hedonic price indices and exploring the many facets of the theory in terms of nonmarket valuation. Freeman (1978), Harrison and Rubinfeld (1978a,b), and Bender, Gronberg, and Hwang (1980) were the first papers to empirically implement the two stage procedure put forth in Rosen (1974) these studies did not exploit the entire system of demand and supply functions, choosing to investigate the properties of demand for air quality only by assuming that the supply of housing was fixed.

Witte, Sumka, and Erekson (1979) were the first to empirically tackle Rosen's two-stage idea on both sides of the market, applying his method to the rental housing market in North Carolina using three stage least squares. They estimated two variants of the model, one that included both product characteristics and supplier/demander characteristics and another that used supplier/demander characteristics as instruments for product characteristics. Their empirical results were quite confirming of Rosen's theory. They found that for the attributes under study, all three showed diminishing marginal returns in the bid function, suggesting concavity, and constant or increasing marginal returns in the offer function, suggesting linearity or convexity, respectively. Rosen intimated that these would be reasonable assumptions for both the structure of costs across producers and the nature of utility across buyers.

Rosen's method also pointed to the downfall of using the single stage hedonic method to perform aggregate benefits analysis. Figure 1.5 shows graphically how interpretations based from the marginal hedonic price function can overstate the true benefits estimates, which come from the MRS function. The true benefit is represented by the area $z'acz''$, while the benefit estimated from the hedonic price function would be $z'abz''$. Thus there is a positive, estimated bias of abc. [8] Harrison and Rubinfeld (1978b) discuss three separate sources of bias when one uses the price index rather than the two-step method of Rosen.

Following in the footsteps of Harrison and Rubinfeld (1978a,b) and Witte et al. (1979), Bloomquist and Worley (1981a,b) also applied Rosen's two-step method, investigating the difference in benefit estimates across functional forms of the initial hedonic price regression as well as those calculated directly from the coefficients from the hedonic regression of step one. They used single market data for their analysis as opposed to the multimarket data approach used in Witte et. al.

[8] However, given that the MRS function does not truly represent a demand curve, benefit calculations may be either over or understated depending upon the marginal utility of income.

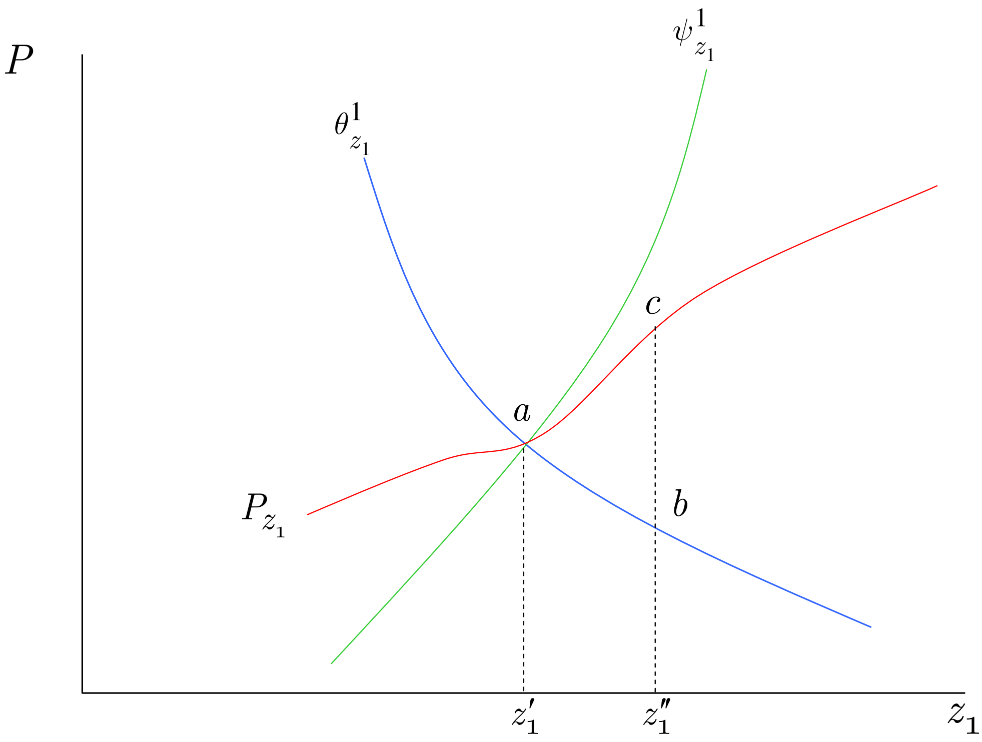

Figure 1.5: Bias in Using Hedonic Price Regression to Calculate Benefits

Palmquist (1984) was one of the first to use local data instead of census tract data used in previous hedonic estimation studies. Thus he was able to gain further insights into the housing market from this disaggregated data. He followed the same approach as Witte et. al. and estimated a separate hedonic price function for each city of study, and then aggregated the marginal prices for the second stage system analysis, thus making the assumption that the demand curves were the same across cities and giving him enough restrictions to identify parameters of interest. One interesting critique of Palmquist's analysis is that his first stage hedonic regressions were linear in variables except for the square footage variable. This was quite different from previous studies as many believed the hedonic price function to be entirely nonlinear. However, later research citing Palmquist has not raised this criticism. Other early applications of Rosen's method were Linneman (1980, 1981) and Bajic (1984).

1.5 Econometric Critiques of Rosen's Two-Step Method

After Rosen proposed the two stage method, and because there was no formal analysis of the econometric implications of it, the first few papers simply treated the system of equations as a standard econometric problem. However, the econometric issues that have been raised against treating the two stage method as a standard system of equations can be summarized into three distinct categories:

1. Identification of structural parameters in the second stage.
2. The nature and causes of endogeneity of product attributes in the system of equations.
3. The choice of functional form for the hedonic price function as well as the marginal rate of substitution and compensated supply functions.

1.5.1 Identification

The arguments behind structural parameter identification are complicated because implicit market prices are not observed, but constructed from some other regression. The fact that these prices have to be estimated causes identification issues that go beyond the standard rank and order conditions inherent in simultaneous equation estimation.

Brown and Rosen (1982) was one of the first paper to seriously take on the econometric implications of Rosen's two stage method. Their note showed that if a researcher specified a quadratic form for the hedonic price function, and then proceeded to specify a linear form in the resulting marginal system analysis (as would be the case for the NQL model), then no structural information would be gained. The intuition behind this result is clear. If one estimates a specified functional form, and then proceeds to take its derivative and estimate the derivative using the same functional form of the derivative then a perfect fit will ensue.

Brown and Harvey Rosen's paper did not criticize Rosen's method per se, merely, brought to the front the way that it was to be estimated econometrically. Their point was that certain restrictions were required to identify the second stage structural demand and supply parameters. In Witte, Sumka and Erekson's case they allowed the hedonic price of rental units to vary by city, but constrained the supply and demand functions to be identical across cities. Harrison and Rubinfeld (1978) also avoided the pitfalls brought up by Brown and Rosen by placing functional form restrictions in their second stage analysis.

In a more detailed analysis Brown (1983) discussed this case in further detail. The conclusions reached by Brown were that the construction of marginal prices, as opposed to outright knowledge of them, was the main reason why their were so many econometric flaws/issues in the two-step method. The main point was that in structural estimation of implicit markets, marginal prices constructed from product attributes must not vary in a collinear fashion with the product attributes which appear in the structural equations. This imposed some functional form restrictions on the hedonic price index as well as the structural equations.[9] Even if the perfect collinearity between marginal prices and right hand side variables is avoided, Brown points out that multicollinearity could still lead to biased estimates of the structural parameters.

One solution to the identification problem that was already in use (although the identification problem was not formally addressed) was multi-market data. Here different hedonic price functions were estimated for separate markets, but the underlying structural demand and supply functions were assumed to be the same. This allowed the marginal prices to vary in such a manner that structural parameters could be uncovered. Even though Brown pointed to the benefits of multi-market data, he was one of the first (in this line of research) to suggest that it was not necessary. The reason that single market data could be used, according to Brown, was that the nonlinear structure of the hedonic price function lead to price variation in implicit markets (nonconstant constructed marginal prices) which was enough to pull

[9]Of course, the hedonic price function could take any form as long as the structural equations are sufficiently restricted to not vary collinearly with the constructed marginal prices.

out the structural parameters of interest, thus alleviating the need for multi-market data, a potential data collection problem.

Another issues raised by Brown was that certain types of restrictions could not be tested given the fact that marginal prices cannot be collinear with right hand side variables. This was a serious drawback in his opinion and one that could potentially be alleviated with cross-market data. Thus, while there was no reason to explicitly state that multimarket data was necessary for structural estimation, in Brown's eyes it lead to more flexibility for the researcher at the cost of more stringent data requirements. One point worth noting is that Brown was clear that multimarket data was no panacea for all of the econometric issues associated with Rosen's two-step method.

The use of multi-market data for structural parameter identification is show graphically in Figures 1.6 and 1.7.

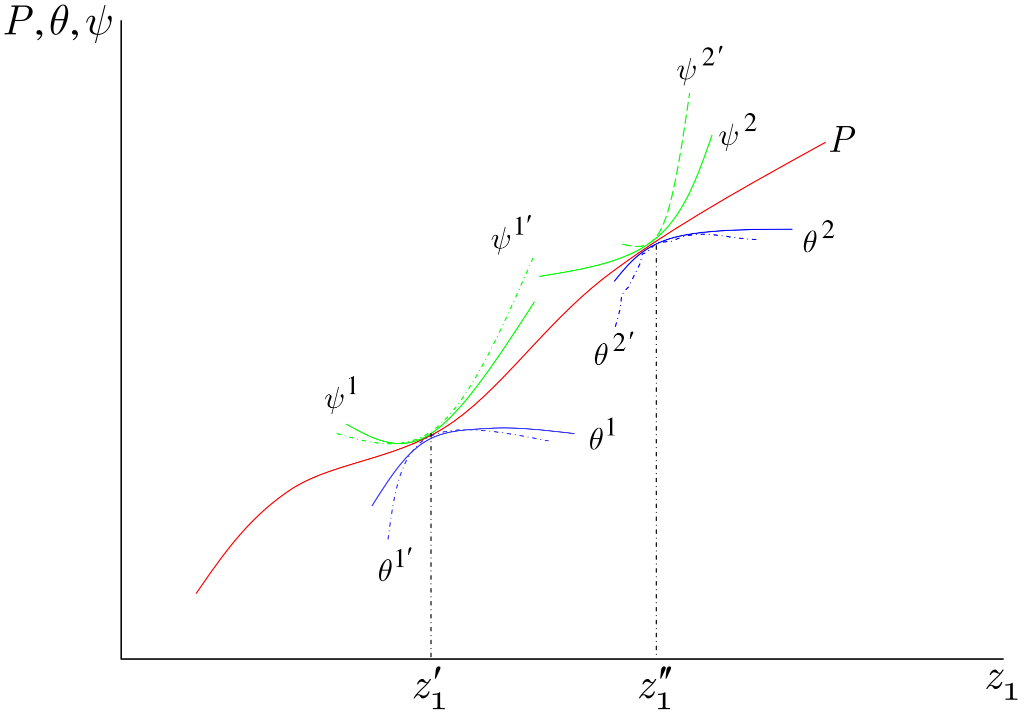

Figure 1.6: Using Multi-Market Data for Structural Identification: Hedonic Equilibrium

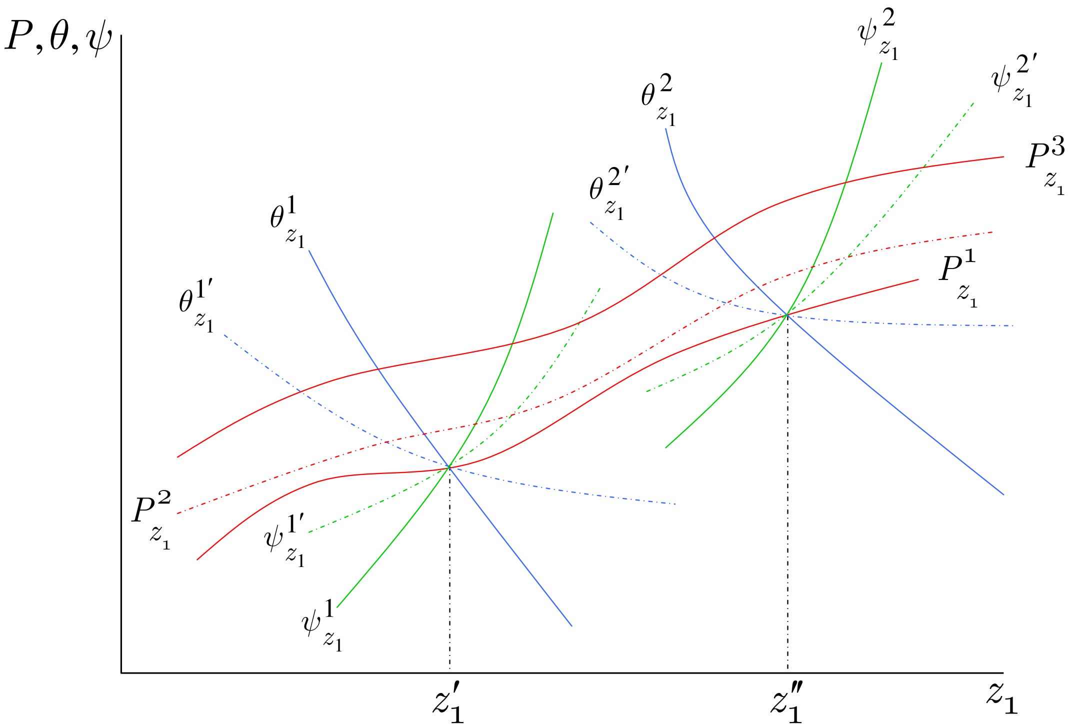

Figure 1.7: Using Multi-Market Data for Structural Identification: Implicit Market Equilibrium

Figure 1.6 shows two different sets of bid and offer curves which trace out the same hedonic price function, with explicit sales at z_1' and z_1''. Moving to the second stage analysis, there will be two demand and two supply curves at the these points, due to the first order conditions, (1.3) and (1.9). Here we see that due to the possibility of more than one demand (supply) curve passing through points z_1' and z_1'', identification of the true demand (supply) curve requires a shift in the marginal price function P_{z_1}, which multimarket data provides. Ohsfeldt and Smith (1985) have examined how much variation must exist between cross market hedonic price gradients using a Monte Carlo analysis, and have found that, at a minimum, price gradients must vary on the order of 20% across markets for the structure of demand (supply) to be identified. While this number may seem high, several studies have reported cross market variation that is inline with the necessary 20%. Another point about the required variation for cross market price gradients was raised by Mendelsohn (1987). He gave no formal amount that gradients must vary but did point out that the elasticities of demand (supply) even in the neighborhood of the equilibrium, were sufficiently similar that cross market differences would not be able to pick them up with any precision. Thus, cross market gradients $P^1_{z_1}$ and $P^2_{z_1}$ will not be enough to uncover the demand (supply) structure, but cross market gradients $P^1_{z_1}$ and $P^3_{z_1}$ will. While the fact that identification of the underlying structure is possible, the requirement of multimarket data is quite burdensome.

Brown and Rosen's paper led to widespread investigation of further econometric implications of Rosen's theory as well as set in motion the search for general conditions under which the second stage parameters could be identified. Mendelsohn (1985) gave some general conditions under which the structural parameters could be identified with single market data. One important conclusion from his paper was that the results of Brown and Rosen did not generalize; they were dependent on arbitrary functional form specifications made in both the first stage and the second stage, something brought up recently by Ekeland, Heckman, and Nesheim (2002, 2004). His idea was to extend the model of Brown and Rosen by allowing the marginal price function to be nonlinear. This added nonlinearity in the second stage proved to be sufficient for identification. Unfortunately, Mendelsohn did not discuss estimation of his method, nor did he include a detailed analysis of how error effects his results. Given this, Bartik (1987a) showed that single market data, in the presence of unobserved tastes, may not be able to identify the underlying structure within the second stage.[10]

McConnell and Phipps (1987) also investigated the issue of preference parameter identification. Their arguments related to the nonlinear structure of the hedonic price function and

[10]Mendelsohn (1987) provides a short survey of the issues with identification of the structural parameters that arose after Rosen's groundbreaking insight into hedonic markets. This survey contains the basic intuition of Brown (1981), Brown and Rosen (1982), Mendelsohn (1985), Ohsfeldt and Smith (1985), and Bartik (1987a).

is something that would arise later on in the literature on identification and estimation.[11] They established criteria when the recovery of information about the preference parameters was obtainable. Their model was set up to be consistent with preference theory and the nonlinear budget constraint is the driver in terms of obtaining identification. One point they raise is that since the budget constraint is nonlinear, the marginal prices of the hedonic price function represent marginal rates of substitution and not Marshallian demand functions, as is common in the presence of utility maximization in the presence of a linear budget. This point was also raised by Murray four years earlier, but they failed to cite him on this point. They make a similar claim as Diamond and Smith that when price taking behavior is present on both sides of the market that no identification problem is present, but when this assumption breaks down identification is suspect because prices and quantities are now jointly dependent.

Their main argument is that the estimation and identification problem is a one step proposition as opposed to Rosen's suggestion of a two step procedure. The reason being is that the marginal price are related to the hedonic price function and so they should be estimated jointly, not one at a time. This claim was the same as Horowitz (1987), though no reference was made to his paper. They laid out conditions where the parameters of the hedonic price function and the marginal rates of substitution could be found. In their setup the hedonic price function is a structural equation in the estimation process and along with the marginal price equations constitute a square system (q unknowns and q equations).

Horowitz did not focus on one specific econometric problem of hedonic equilibrium, proposing four different areas that needed attention. His four main points were that (i) explicit specification of the underlying market functions (utility and cost) may not always be able to solve the identification problem raised by Brown and Rosen, (ii) except in special cases, the stochastic term in the marginal cost and marginal rate of substitution functions would result in equations that would not have a normal regression form, (iii) only in excep-

[11] See Ekeland, Heckman, and Nesheim (2002, 2004) for a detailed discussion of identification through nonlinearity.

tional cases would the attributes in the first stage hedonic price regression not be endogenous, and (iv) if certain utility functions (strongly separable in attributes) are pre-specified a powerful specification test can be constructed. His research has been cited from time to time but many have avoided following (ii) and (iii) as these points are tough to work with empirically.

1.5.2 Simultaneity/Endogeneity

Brown (1981), Murray (1983), and Diamond and Smith (1985) were among the first to raise the issue of simultaneity between constructed marginal prices and observed attribute levels. Brown, however, only briefly discussed this point, choosing to focus more on identification and estimation of structural parameters, ignoring the possible endogeneity of product attributes.

The uncovering of the simultaneity in the hedonic system was important because the answer to that question was key in determining which variables can be used as instruments and how the demand system should be estimated to obtain consistent estimates. Diamond and Smith were perhaps the first to address the nature of the simultaneity when both firms and households are price takers. Their model formulation then could assess the type of data needed as well as the correct estimation procedure if one was interested in say, the demand for air quality. While their analysis was geared specifically toward the housing market, they were able to show that when price taking behavior is indicative on both sides of the market then there is no interface between demand and supply, in terms of individual decisions. What they did uncover though was that simultaneity did exist in the estimation, due to the nonlinear structure of the hedonic price function. This was key in their formulation as the choice of a bundle of characteristics (to buy or sell) also meant the simultaneous choice of a marginal price for each of the characteristics. They also showed that since supply shifts simply resulted in movements from one demand curve to another, supplier characteristics were not powerful instruments in the demand system, contrary to Rosen's suggestion of using demand characteristics as instruments in the supply equations and supplier characteristics in the demand equations. A graphical depiction of attributes being endogenous in the second stage estimation is shown in Figure 1.8.

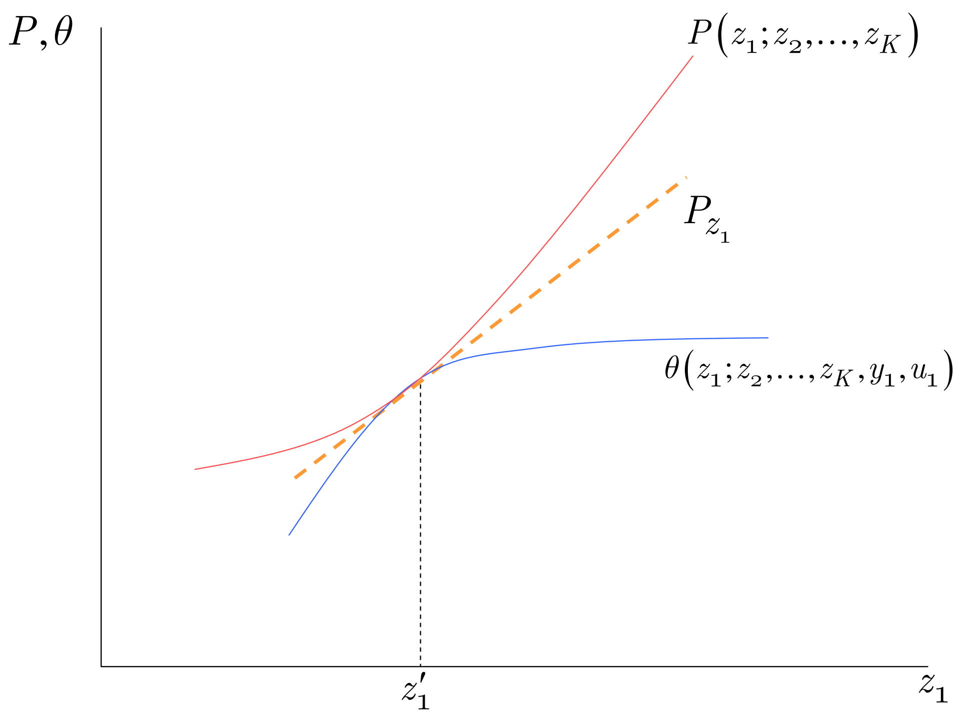

Figure 1.8: Endogeneity of Attributes in the Second Sage

The simultaneity of buyers and sellers pointed out by Rosen for the hedonic price function was revisited by Diamond and Smith (1985) investigating what information was present in the demand and supply equations of the second stage system. Thy claimed, at least for the housing market, if both buyers and sellers are acting as price takers then neither has an impact on the other. In other words, if prices are fixed then so are marginal prices and so any shift in demand results in a buyer moving to a different supply curve (not along a supply curve) and vice versa. This is depicted in Figure 1.9.

Here, a shift in a buyer's MRS function from $\theta^1_{z_1}$ to $\theta^{1'}_{z_1}$ does not represent a movement along the corresponding seller's supply function, $\psi^1_{z_1}$ in equilibrium. Due to the price taking behavior of agents on both sides of the market, buyers treat the marginal hedonic price

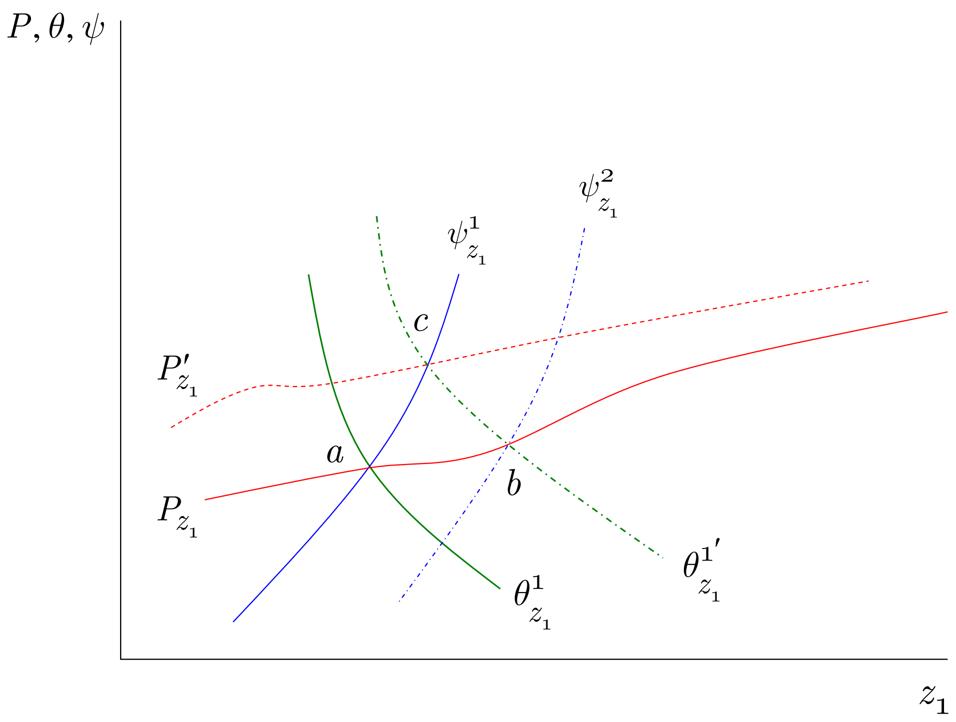

Figure 1.9: The Simultaneity Issue in Implicit Markets

function as the supply schedule while sellers treat it as the demand schedule. Thus, any shift in the MRS function requires the buyer to locate at a point on the marginal hedonic price function consistent with a different seller, $\psi^2_{z_1}$. So, if a consumer is originally in equilibrium at a and a demand shock occurs, the new equilibrium is at b instead of c, as would be the case in a standard market. Point c would represent an equilibrium if the marginal price function were P'_{z_1} instead of P_{z_1}. Thus, MRS shifters cannot be used in the second stage estimation to determine the shape of the supply functions nor can supply shifters be used to uncover the structure of MRS functions in implicit markets. Also, due to the nonlinearity of the hedonic price function, the marginal price function will be nonconstant, implying that it is not perfectly elastic and so quantity and the corresponding marginal price are determined simultaneously, as was portrayed in Figure 1.8.

So, the time frame of the market is irrelevant. No matter if the market is in the short run or the long run the buyer problem can be treated separately from the seller problem because their actions are independent form one another.[12] Their implication was that the entire system of demand and supply did not need to be estimated at the same time, they could be thought of as separate events. Even though they showed that there is a disconnect between buyers and sellers in the second stage it still did not rule out endogeneity of the characteristics, that still remains so long as the hedonic price function is nonlinear.

Aside from the identification problems introduced by Brown and Rosen, both Bartik (1987) and Epple (1987) discussed the issue of endogeneity.[13] Their argument goes as follows. Given that buyers' decisions do not affect how suppliers provide the good in question, both the quantity of particular characteristics (typically the right hand side or exogenous variable) as well as the slope of the hedonic price function (the variable of interest in the second stage analysis) are deemed endogenous because buyers select both at the same time. That is, by selecting a given quantity of a characteristic, implicitly a marginal price is also being selected. Thus, an estimation technique such as two or three stage least squares where endogeneity is not controlled for will yield biased slope coefficients and introduces another problem with consistently uncovering structural parameters.

Although Rosen discusses identification problems in the estimation of his system of compensated demand and supply curves, he does not propose a formal econometric technique, merely that system estimators be used. However, if one were to use two-stage or three-stage least squares, then the suitable instruments may not be valid, as point our by Bartik. Suppose that one is interested in a given demand equation, then the argument would be that using the characteristics of the suppliers could serve as instruments for the levels of the qualities observed. Bartik considers the housing market and his argument is as follows. Suppose some

[12]Unless of course there are structural errors which are correlated across buyers and sellers.
[13]Kahn and Lang (1988) also raised the same issues as Bartik and Epple.

buyers have a taste for good craftsmanship and so will locate at houses made or lived in by carpenters. Thus, using the fact that the seller is a carpenter (as an instrument), is implicitly linked to the buyer's unobserved taste for good work, which is in turn linked to the quantity of the characteristic craftsmanship where he/she locates when making a purchase.

Bartik goes on to show that anything that shifts the budget constraint will serve as a valid instrument in the demand equations. We does not consider estimation of the compensated supply equation in his analysis. However, continuing with his intuition, it would seem obvious that anything that shifts the hedonic price function, from the sellers standpoint, would serve as a valid instrument when estimating the supply equations. Therefore, time or city dummies may be considered as viable instruments for the quantities of the characteristics on the right hand side of the supply equations.

Epple's paper investigates the same issues as Bartik, but in a much more general framework. His primary tool is the NQL model. Using this particular model, Epple shows that:

1. Regardless of the endogeneity of supply, OLS applied to the demand side of the hedonic market will not yield consistent estimates.

2. OLS applied to the hedonic price equation itself is consistent only if the error term of that equation is uncorrelated to the error terms of the demand side equations.

3. One cannot assume that all supplier and demander characteristics are exogenous.

At first glance, the methodology of Epple seems to doom empirical investigation of markets via the hedonic method.

The reason that Epple's claim hold true is that in the second stage analysis, the marginal prices (the derivative of the hedonic price function) are estimated rather than known. Given that the dependent variable in the second stage is formed from estimates rather than found explicitly, the variables on the right hand side are forced to be correlated with the errors, which gives the inconsistency results expounded by Epple and Bartik. Another point is that because the hedonic price function itself is nonlinear, when consumers choose a quantity to purchase, even though price is exogenously given, their selection is also an implicit selection of the marginal price. That being said, both marginal price and quantity of characteristics are endogenous to the consumer, thus instruments must be used when one of the two is on the right hand side.

Fortunately, a path towards consistent estimation is laid out for the Normal-Quadratic-Linear model. Epple proposes a set of rank conditions which the error terms of the supply and demand side equations must satisfy. Within his derivation of this rank condition, Epple comes across a surprising result which confirms his and Bartik's intuitions, the model specifications in which errors arise only from unobserved characteristics for economic agents and the characteristics of the good in question are not viable. Thus, there must exist measured variables that are correlated with the unobserved terms in the demand and supply system!

For identification, order conditions are also laid out both for the supply side as well as the demand side. Given that there are q variables of interest (the q product characteristics) and $2q$ equations (q demand equations and q supply equations), the entire system is overidentified and conditions are required to obtain the parameters of interest. In Witte et. al. (1979) identification is achieved by assuming that demand and supply functions are the same across markets, this limits the number of parameters to be estimated and identification follows. Harrison and Rubinfeld (1978a,b) were able to identify their parameters of interest, the coefficients in the demand for clean air, using nonlinear functional form restrictions as well as exclusions restrictions on the demand equation.

Epple also brought up the possibility of nonlinear demand and supply functions and the potential restrictions that would be required for consistent estimation and identification in that setting. However, his discussion was quite short and no formal conditions were laid out. Aside from Rosen, Epple was also one of the first to suggest that hedonic models be considered when incomplete or asymmetric information was present as well as looking at dynamic aspects of hedonic markets.

More recently, Cheshire and Sheppard (1998) use instruments that were mentioned by Murray (1983, pg. 330) and have roots in the time series literature where lagged values are used to instrument for current values. Their suggestion is to use 'representative' agent characteristics and their attributes purchased to instrument for the true characteristics and attributes purchased by any given agent. In their paper, 'representative' is chosen as the two agents within some distance or metric from the agent in question.[14] They claim that this method will work extremely well when the data set is void of many agent characteristics over which to search for credible instruments.

1.5.3 Choice of Functional Form

Another obvious, but critical, implication of the two-step method was proper specification of the hedonic price function. Indeed, for structural estimation in implicit markets correct specification is paramount. The reason being that because marginal prices are required to estimate structural parameters, and because marginal prices are constructed from the hedonic price function, any specification error in the hedonic price function will carry over to the second stage analysis and cast doubt on parameter estimates. So, even though misspecification has serious consequences for any econometric investigation, it is doubly troublesome for hedonic markets, given the nature of implicit prices. Brown (1983) goes through a slew of examples that look at what happens when hedonic models (both first and second stage) are overspecified and underspecified to draw conclusions as to the consequences of over controlling for and omitting variables. This fact about two-stage hedonic price estimation turns out to be one of the most popularly studied in investigations of the hedonic price method. Indeed, as with other two-stage methods, the results of the second stage are influenced much more when the stages are linked as opposed to when they are independent.[15]

[14]They scaled all of their data so that it had zero mean and unit variance and used the Euclidean metric.

[15]See Wang and Schmidt (2002) for another type of two-stage estimation that does not work properly.

Aside from the criticisms of the quadratic-linear specification of the two-stage setup raised by Brown and Rosen, several researchers started to investigate the role that functional form played in hedonic price estimation. The first to formally compare results based on flexible functional forms was Bender et. al. (1980), employing the quadratic Box-Cox method proposed in Halvorsen and Pollakowski (1981). Their results showed that traditional nonlinear estimation methods such as linear Box-Cox, semilog, and log-log models were not capable of uncovering the true nonlinear structure of the hedonic price function, nor the underlying marginal rate of substitution functions associated with the market.[16] The fact that the hedonic price function was nonlinear allows the researcher considerable flexibility in modeling strategies and so care must be taken in model specification. Halvorsen and Pollakowski (1981) introduced a hierarchical model that contained as specialized cases many of the simple nonlinear models that had been used up to that point. Their quadratic Box-Cox model contained the linear, quadratic, semi-log, translog, square root quadratic, generalized square root quadratic, as well as homogeneous versions of the linear and quadratic generalized Leontief forms. Indeed their model covered quite a few of the popular nonlinear models. They empirically tested their quadratic Box-Cox against all of the specialized cases for a housing market in San Francisco and rejected all of the alternative, simpler specifications, thus lending credibility to the intuition that model selection was paramount in constructing hedonic price indices.

Not long after Halvorsen and Pollakowski raised the issue of model specification, Quigley (1982) investigated how assumptions on consumer preferences impacted the form of the hedonic price function and the role that a priori specification of the hedonic price function led to uncovering information about the underlying utility functions. To both questions he answered that it does not cause considerable restrictions, which was positive towards specification and estimation of hedonic price systems. Indeed, his results found that utility maximization based on price taking behavior was consistent with a concave hedonic price

[16] This is not entirely true as they only estimated the MRS for air quality, ignoring the remaining MRS schedules. Whether their results would have extended to the other attributes is unknown.

function and so even very restrictive assumptions about utility still led to a general form for a market's hedonic price function. His second result was perhaps even more encouraging, if the hedonic price was completely known, then, given that the budget constraint was nonlinear, the shape of the utility function could be inferred from the slope of the hedonic price function. These two results were used to determine the impact that a housing subsidy program had on low income locals in the city of Santa Ana, El Salvador. Quigley determined that the compensating variation from the program was substantial, an average benefit per household of 530 to 640 units of the local currency.

Quigley's model was based off of a linear Box-Cox functional form of the hedonic price function, with second stage identification coming via explicit specification of the utility function. This was the first paper to estimate the two-step hedonic model by assuming an explicit form of the utility function. The utility function was assumed to be of the generalized constant elasticity of substitution form (GCES). Kanemoto and Nakamura (1986) did not specify the form of the utility function, but suggested specification of the bid functions of consumers. Their quadratic specification of the bid function resulted in an additive, linear in attributes error specification when there were missing attributes, a fact that was not formally introduced in Quigley's model. As Kanemoto and Nakamura (pg. 225) put it "These restrictions imposed on the shape of the bid price function are crucial for identification." They estimated their model as well as Quigley's model and compared results, finding that there were differences in the estimated marginal rates of substitution between the living area of the house and the numeraire good. Their results also suggested that along a ray from the origin, the slopes of the utility function were nonconstant, implying that utility was nonhomothetic, contrary to Quigley's assumption.

In a comment to Halvorsen and Pollakowski, Cassel and Mendelsohn (1985) pointed out that even though flexible functional forms may fit hedonic data better, care must be taken with selecting an appropriate model based on what the researcher is trying to get from the data. Their point was that if one was interested in the marginal effect of a particular variable, then Box-Cox transformations may not be optimal. Or if one was to use their model to forecast prices, a nonlinear transformation with great in-sample fit may actually have poor out of sample fit since the expected value of a nonlinear function is not equivalent to the value of a nonlinear function at the mean of the data. Another problem with the use of Box-Cox methods, linear or quadratic, is the bias in the variance estimates. This point was raised, with hedonic models in mind, by Blackley, Follain, and Ondrich (1984). They showed that t-statistics calculated from the Box-Cox method, not accounting for the bias raised by Spitzer (1982), was a large as 300% in one case and 578% in another case. The magnitude of these differences in significance ratios could lead one to erroneously exclude relevant variables and/or include irrelevant variables.

While their point was very intuitive, Cropper, Deck, and McConnell (1988), using simulated data, found that the standard Box-Cox or the quadratic Box-Cox proposed by Halvorsen and Pollakowski had the best determination of the marginal prices when all the attributes were known, thus refuting one of the main points of Cassel and Mendelsohn. Even when their were missing attributes or proxies for attributes, the linear Box-Cox was found to reasonably determine the marginal prices of the hedonic function. However, several criticisms of the Box-Cox method remain. The Box-Cox transformation cannot be used with negative valued data, specifying one transformation parameter for every variable is has not formal basis, and specifying a different transformation parameter for each variable is highly computationally intensive. Another point worth making is that Box and Cox (1964) noted themselves that the use of the transformation was to make the regressors independent of one another and so use of the quadratic Box-Cox is strange from this standpoint.

Stock (1991) and Meese and Wallace (1991) were some of the first to investigate the benefits of using nonparametric methods to estimate hedonic price indices, although Anglin and Gençay (1996) is the more widely cited paper regarding functional form relaxation in regards to hedonic price indices. Aside from these papers there has been a litany of research investigating the ability of advanced statistical estimation techniques to model hedonic prices.

Table 1.1 contains a list of papers that have relaxed the functional form of the hedonic price function. The columns of the table contain the authors of the paper, the estimation procedure of the hedonic price function, and whether or not the authors carried out the two-stage analysis.

Table 1.1: Categorization of Specification/Functional Form Relaxation Analyses

Authors	Estimation Procedure	Two-Stage Analysis? (Yes/No)
Bender et. al. (1980)	Quadratic Box-Cox	Yes
Atkinson and Crocker (1987)	Bayesian Covariate Selection	No
Cropper et. al. (1988)	Quadratic Box-Cox	No
Graves et. al. (1988)	Bayesian Covariate Selection	Yes
Rasmussen and Zuehlke (1990)	Quadratic Box-Cox	No
Hartog and Bierens (1991)	Nonparametric	No
Meese and Wallace (1991)	Nonparametric	No
Stock (1991)	Semiparametric	No
Can (1992)	Spatial Expansion	No
Coulson (1992)	Semiparametric	No
Pace (1993)	Nonparametric	No
Pace (1995)	Nonparametric	No
Knight et. al. (1995)	Varying Coefficients	No
Anglin and Gençay (1996)	Semiparametric	No
Gençay and Yang (1996)	Semiparametric	No
Mason and Quigley (1996)	Nonparametric	No
Wallace (1996)	Nonparametric	No
Meese and Wallace (1997)	Nonparametric	No
Pace (1998)	Additive Semiparametric	No
Iwata et al. (2000)	Additive Nonparametric	No
Lee et. al. (2000)	Average Derivative Estimation	No
McMillen and Thorsnes (2000)	Semiparametric	No
Clapp et. al. (2002)	Bayesian Smoothing	No
Bao et. al. (2004)	Spline Smoothing	No
Bin (2004)	Additive Semiparametric	No
Clapp (2004)	Semiparametric	No
Bajari and Kahn (2005)	Nonparametric	Yes
Bin (2005)	Additive Semiparametric	No
Martins-Filho and Bin (2005)	Additive Nonparametric	No

One common theme of the papers is that they consider only a first stage analysis and so the effects of hedonic price misspecification on the second stage results has gone unstudied. Another caveat about these first stage methods. Many nonparametric methods, until recently, could only model continuous variables within an unknown function. This downside required modelers to enter discrete attributes in an additively separable, linear fashion. This has two distinct consequences. The first is that the unbundling of the good assumption does not correspond to constant marginal prices (adding the discrete variables in a linear fashion) while the second is the implicit assumption that the derivatives of the hedonic price function in the continuous variables (the marginal prices) is independent of the discrete variables. Whether these assumptions are valid or not depends upon the good in question and the market being used for the study, but the consequences of these facts need further attention.

1.6 Some Recent Econometric Insights Into Rosen's Two-Step Method

Recently, Ekeland, Heckman, and Neshiem (2002, 2004) (EHN hereafter) laid out a potent mathematical formulation of the economics underlying hedonic equilibrium that addresses the main econometric criticisms of his proposed two-stage approach. They begin by showing that the NQL model, the primary model used in the late 70's to employ Rosen's method and which the brunt of the criticisms has been levied, is non-generic. A parameter or model that is non-generic in laymen's terms means exceptional. Mathematically what we see is that small perturbations of any of the assumptions surrounding the NQL (either normality or any of the functional forms) will lead to a model that does not closely resemble the NQL and will be highly nonlinear in both the hedonic price function as well as in the marginal price function. EHN give an explicit example where the functional forms are quadratic, but they use mixtures of normals. Even with mixing weights of 0.999 and 0.001 the resulting marginal hedonic price is highly nonlinear.[17] Intuitively, since the underlying economic structure is

[17]See Figure 1 on page 308 of EHN (2002).

rarely observed, what should be studied are the properties that are typical (referred to as generic in the topographic literature) and do not occur only in rare or special situations.

A property of an economic phenomena is typical or 'generic' if it displays stability under perturbations to the economic structure. Thus one would expect that in a world where little is witnessed, those properties which are typical are the most likely to be seen. This is what EHN prove about the NQL model: that it is not a typical structure for implicit markets and so the possibility of observing it, while not zero, is highly unlikely. Indeed, with no other restrictions the NQL allows for negative quantities of attributes to be bought and sold as well as negative prices to occur. What this point emphasizes is that the claim of Brown and Rosen of not being able to identify anything in the second step without arbitrarily assumed functional forms is not intrinsic to the hedonic setup, but to the capriciousness of the modeling strategy; the collinearity induced because of the assumed linearity, goes away when one resorts to nonlinear/nonparametric methods.

Mathematically, a generic property means that an open, dense subset exists for the topological property space. That is, for a property[18] F, if there exists a sequence of subsets in the parameter space Θ, $U_r \subset \Theta$ such that U_r is open and dense (also known as residual) $\forall r$, and $A(F) \supset \bigcap_r U_r$, then the property is generic. The property occupies the set

$$A(F) = \{\theta | F(\theta) \text{is true}\}.$$

A property that holds for any parameter values is defined as $A(F) = \Theta$. In economics these properties are rare. What is more common are properties that hold except under exceptional circumstances. These kinds of properties are generic.

In our case the parameter space would be the distribution of tastes across buyers, the distribution of technology across sellers, the functional form of the utility function and the functional form of the cost function.[19] Our property is the identification of the structural parameters in the implicit markets. As Brown and Rosen suggested over 20 years ago, without

[18] We use the nomenclature F for feature, since P is used to reference price.

[19] The NQL model parameterizes all distributions to be normal and all functional forms to be quadratic.

multimarket data the structure of the implicit markets is not identified in the NQL model. EHN show that this parameterization is nongeneric, meaning that a small perturbation of either normality or the quadratic functional forms lies in an open dense subset of the parameter space. This is shown pictorially in Figure 1.10. There the grey area is the part of the parameter space where the implicit market structure is identified. We see that within any neighborhood of the NQL model the structure is identified. Small perturbations take us from the white area to the grey area. Thus, the NQL model is not typical of the underlying hedonic structure.

$$A(F) = \{\theta \mid F(\theta) \text{ Identification}\}$$

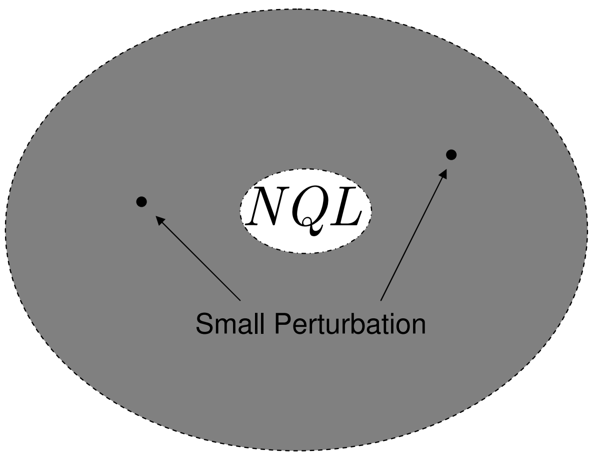

Figure 1.10: The Nongenericity of the Normal-Quadratic-Linear Model

EHN show how to use the nonlinearity of the system of attributes to acquire identification of the parameters of interest as well as to specify good instruments. Their point about endogeneity is that while exclusions restrictions do not exists, there do exist nonlinear transformations that can act as valid instruments. Another of their important insights is that the Rosen method will work even for data on a single market. There claim is that it makes **no sense** to assume that preferences, costs, and demand and supply behavior are the same across markets while hedonic price functions are different. Thus, they show that having to resort to those type of restrictive assumptions are not necessary for identification or controlling for endogeneity. Again, EHN (2004) set up, for the additive, scalar attribute case, the econometric model as a two stage procedure, ignoring the point of Horowitz and McConnell and Phipps that the attributes may be endogenous in the hedonic price regression function. Conditions for consistent and efficient estimation of both the first and second stage are laid out although no formal application is provided. Heckman, Matzkin, and Nesheim (2003) discuss the nonadditive, scalar attribute case and Heckman, Matzkin, and Nesheim (2002) lay out a procedure for the nonseparable, vector attribute framework.

Another recent advance in the estimation of hedonic price index estimation, especially in the housing market, has been the use of spatial methods. For example, there may be a correlation between selling prices that has nothing to do with attributes *per se*, but with the fact that a previously sold house was sold above market value, pushing future home sellers, within the same neighborhood, to arbitrarily list their houses for higher than market values. This effect shows up in the residuals as a spatial correlation and needs to be taken account of to obtain consistent estimates of the hedonic attribute coefficients. Another source of spatial correlation may be an unobserved geographical attribute such as *view* or *waterfront*. These variables inevitably raise the price of houses that possess them, but if uncontrolled for will cause the errors of the hedonic price function to display cross sectional correlation, something typically assumed away in cross sectional econometric analyses. An excellent compendium on spatial methods, with specific reference to hedonic methods is the edited volume by Lesage and Pace (2004).

Other areas that have received burgeoning attention with respect to the estimation and understanding of hedonic price models has been the incorporation of bargaining and search within the modeling process. In Rosen's initial methodology, everyone is fully informed and markets are perfectly competitive so that search and bargaining are irrelevant in the hedonic framework. However, there have been many attempts to model the impact that either bargaining, search, or both have on the hedonic price index. Once again, these investigations have focused on the first stage analysis only leaving the impact on the second stage results as an area for future research.

The main idea behind bargaining/search in implicit markets is that because each good is a bundle of attributes, each good is unique[20] and so there is no well defined market, which makes bargaining a relevant fact of implicit markets since there is no obvious intuitive reason that bid and offer curves should align and be tangential to one another, creating a smooth, well-defined hedonic price function. Also, due to the supposed uniqueness of goods, there is no reason to believe that there are enough sellers and buyers that markets will be competitive and price takers. This invokes the need to incorporate search into the modeling framework as both buyers and sellers can shop around or hold out, respectively for better offers, thus going against Rosen's original methodology.

1.7 Conclusions and Areas for Future Research

While many advances have been made since the first few papers empirically implemented Rosen's two-stage methodology, much more remains. The housing market has received the most empirical consideration, but many other goods have been investigated. These include but are not limited to prostitutes, used coins, cars, newspapers, and wine. An even wider array of goods should be considered to make the method more appealing to economists outside of urban economics and regional studies. As with any other subfield of economics ,the research that is being done is constantly changing the scope and direction of other's thoughts and it is the aim of this and subsequent chapters to help the focus of those interested

[20]Some authors have called these markets 'thin', implying that there are so few (or one) good(s) on it that not all demand is satisfied or enough supply. An example would be a Picasso painting.

in understanding and applying hedonic methods both for single and two-stage analysis of implicit markets.

The literature investigating hedonic price index construction and the uncovering of structural parameters is vast, but there remains much to do. The use of fully nonparametric methods, not just semiparametric methods as a viable estimation technique needs to be taken into account. The role of information, search, and bargaining need to be considered in greater depth both theoretically and empirically. The role that producer and buyer characteristics play in a first stage analysis need to be considered in more detail and estimation methods for discrete attributes needs greater attention. More attention between nonparametric methods and spatial dependence also needs further investigation. The following chapters will investigate the impact of not fully relaxing functional forms and the consequences of the failure to do so, as well as providing a formal modeling strategy for the incorporation of search and less than full information on both sides of the market and the effects this has on estimation and interpretation of hedonic price indices.

Chapter 2

Fully Nonparametric Estimation of an Hedonic Price Index[1]

2.1 Introduction

Sherwin Rosen's (1974) theory of hedonic prices led to a proliferation of research in the valuation of all varieties of economic goods. One of the major focuses of applied hedonic investigations has been in the housing market. Following Rosen's idea that "... it is inappropriate to place too many restrictions on [the hedonic price] at the outset," many applied papers have tried to relax functional form restrictions to facilitate more realistic marginal valuations (determine implicit prices) of housing characteristics. An exemplar typifying this approach was considered by Anglin and Gençay (1996), hereafter AG, who used a semiparametric model to determine implicit prices of housing attributes. We compare our findings from a fully nonparametric model with the results from AG. We find that our nonparametric model leads to a richer analysis of implicit prices by allowing a more complex underlying structure than the semiparametric specification of AG.

Given that differentiated products cannot be unbundled, the hedonic price function has a nonlinear structure. However, for attributes that are dichotomous or discrete (such as having central air conditioning and the number of bathrooms in a house), flexible functional forms as well as semiparametric and most nonparametric methods do not adequately model these types of attributes, resorting to include them in a linear fashion. This inability of advanced estimation methods to account for the potential nonlinearity of dichotomous and/or polychotomous attributes within the hedonic framework is a serious downside to employing them,

[1] An abbreviated version of this chapter appears as Parmeter, Henderson, and Kumbhakar (forthcoming).

given that model misspecification can have serious consequences in any second stage analysis (see Brown (1981) and Palmquist (1987)). Recent insights into modeling discrete variables in a nonparametric framework (see Li and Racine (2004) and Racine and Li (2004))are thus well suited to the estimation of differentiated products when some or all of the product's attributes are (di)polychotomous.

As discussed in Chapter 1, there have been many papers that have attempted to relax the functional form of the hedonic price index in the first stage. Table 1.1 lists the major contributions to this literature. Almost all of those papers either treat discrete variables in a linear fashion, or they use only continuous variables. In both of these scenarios, the model is highly unlikely to do an adequate job of describing the underlying structure of the hedonic market if these variables either belong in the model and are erroneously omitted or enter in a nonlinear fashion.

In hedonic price models it is argued that the value of a good (a house in the present case) depends on the amounts of attributes it contains. Thus, the good's price will be a function of its attributes/characteristics. Implicit prices of the characteristics can be computed from the partial derivatives of the price function. Since these derivatives may be dependent upon the levels of these characteristics, the choice of functional form in an empirical analysis is quite important. Thus, econometric techniques such as nonparametric and semiparametric methods which make few or no restrictions on the functional form of the hedonic price model should provide more reliable information about implicit prices. However, because semiparametric methods typically model discrete variables in a linear fashion, this chapter aims to discover the interpretive differences yielded from the standard semiparametric method, versus nonparametric generalized kernel estimation.

The remainder of the chapter is as follows. Section 2.2 discusses the benchmark semiparametric method employed in the literature today and also reviews in detail the nonparametric generalized kernel regression framework as well as cross-validatory techniques for both estimation procedures. Section 2.3 presents a comparative analysis between the two methods to see what insights the generalized kernel method provides. Section 2.4 concludes and offers some proposals for future research using this new method.

2.2 Nonparametric versus Semiparametric Estimation

From Rosen's theory the hedonic model can be written as

$$P = h(z, \gamma) + \varepsilon^h. \tag{2.1}$$

Here P is a $n \times 1$ vector of prices for the product, z is the $n \times q$ matrix of characteristics, γ is a vector of parameters associated with the hedonic price function, and ε^h is an error term corresponding to measurement error of the attributes as well as accounting for unobserved (by the econometrician) characteristics.[2] In this setup there are n observations. Modeling the hedonic price function typically begins with selecting a functional form for $h(z, \gamma)$ and proceeding from there. The next two subsections discuss estimation of $h(z, \gamma)$ when either all or part of the function is unknown.

2.2.1 Semiparametric Estimation

Semiparametric models were first studied (in the econometrics literature) by Robinson (1988) and Stock (1989)[3] to take advantage of a known parametric relationship and an unknown functional relationship in economic models. A semiparametric model for estimating a hedonic price function takes the form:

$$P = z_1\beta + h(z_2) + \varepsilon^h. \tag{2.2}$$

In this setup, z_1 is a $n \times q_1$ matrix of attributes entering linearly, β is a $q_1 \times 1$-dimensional vector of constant marginal effects, while z_2 is a q_2-dimensional vector of attributes that

[2]Errors in optimizing may also arise, however, as Horowitz (1987) points out, this may lead to a situation where the estimating equation cannot be written in regression form. We ignore errors in optimizing in what follows.

[3]Stock (1991) was the first to estimate a hedonic price function using a semiparametric specification, however, he was only concerned with the first stage and conducted his benefits analysis from these results. The criticisms against the use of the one-step hedonic method to calculate benefits are valid here as well even though the hedonic price function is estimated semiparametrically.

effect the dependent variable through the unknown function $h(\cdot)$, where $q_1 + q_2 = q$. In previous studies using semiparametric methods to estimate hedonic price functions, z_1 is usually composed of discrete and dummy characteristics while z_2 represents the continuous characteristics of the good in question.

Semiparametric models must be estimated in two steps due to the interplay between the parametric and unknown portion. We first estimate the parametric component by eliminating the unknown portion of the hedonic price function. This is done by taking conditional expectations of P and z_1 with respect to z_2. Let $E_P = E[P \mid z_2]$ and $E_{z_1} = E[z_1 \mid z_2]$ denote the appropriate conditional expectations. Assuming $E[\varepsilon^h \mid z_2] = 0$, the conditional expectation of Equation (2.2) can be defined as

$$E_P \equiv E_{z_1}'\beta + h(z_2). \tag{2.3}$$

Subtracting (2.3) from (2.2) yields

$$P - E_P = (z_1 - E_{z_1})'\,\beta + \varepsilon^h. \tag{2.4}$$

Equation (2.4) can be estimated by OLS once suitable estimates have been found for E_P and E_{z_1}. These conditional expectations can be found using standard kernel smoothing methods, which are described in detail next.

Kernel smoothing methods (for continuous variables) are a variant of weighted least squares, with the weights being proportional to the distance between evaluation points and data points. Thus, data points closer to the evaluation point should receive more weight than points farther away. The weighting involves two components, a weighting function, typically referred to as a kernel, and a measure of distance between points, typically referred to as the bandwidth. Essentially, the kernel function places a large weight to a point close to the evaluation point and a small weight to a point far away from the evaluation point. Here "close" and "far" are determined by the bandwidth. When the bandwidth is small even small differences between points is magnified, resulting in smaller weights while for large bandwidths larger differences between points are pared down resulting in larger weights.

The difference between a large bandwidth and a small bandwidth graphically is a function that looks relatively smooth and one that is choppy with many cusps, respectively. Given that the bandwidth is commonly associated with the smoothness of the function we typically refer to bandwidths as smoothing factors. A standard kernel smoothing technique employed for estimating a conditional expectation is the Nadaraya-Watson, or local-constant-least-squares (LCLS), regression estimator. For a given evaluation point, z_2^o our conditional expectations may be written as

$$E_P = \frac{\sum_{i=1}^{n} K_i(z_2^o) P_i}{\sum_{i=1}^{n} K_i(z_2^o)}$$

$$E_{z_1} = \frac{\sum_{i=1}^{n} K_i(z_2^o) z_{1,i}}{\sum_{i=1}^{n} K_i(z_2^o)},$$

where $K_i(z_2^o) = K\left(z_{2,i}, z_2^o, \omega\right)$ is the kernel function and ω is the bandwidth.[4] Once a kernel function and bandwidth have been selected, the conditional expectations can be estimated and the parameters of (2.4) can be estimated via OLS. To stress the dependence of the slope coefficient estimates on the choice of bandwidth we will write $\hat{\beta}(\omega)$ in all formulas that follow. Usually, the evaluation points are just the data points taken one-by-one, so for n observations there would be n evaluation points. One of the most popular choices of the kernel function (in econometric applications) is the Gaussian product kernel (see Pagan and Ullah (1999)). That is,

$$K_i(z_2^o) = \prod_{j=1}^{q_2} k\left(\frac{z_{2,ij} - z_{2,j}^o}{\omega}\right), \tag{2.5}$$

where $k(\varsigma) = \frac{1}{\sqrt{2\pi\omega}} e^{-.5\varsigma^2}$. To obtain an estimate for the unknown function, $h(z_2)$, one would use the corresponding estimates of β from (2.4) in Equation (2.2) to obtain,

$$P - z_1' \hat{\beta}(\omega) = h(z_2) + \varepsilon_{uo}^h. \tag{2.6}$$

[4] The kernel function, $K_i(\cdot)$, is typically assumed to be symmetric for technical purposes relating to the bias and variance of the estimator.

Again, using the fact that $E[\varepsilon^h \mid z_2] = 0$, we have that $h(z_2) = E[P - z_1'\hat{\beta}(\omega) \mid z_2]$, which can be estimated similarly to E_P and E_{z_1}. Denoting $\eta_1 \equiv P - z_1'\hat{\beta}(\omega)$, the estimate of the unknown function in Equation (2.6 becomes

$$E_{\eta_1} = \hat{h}(z_2^o) = \frac{\sum_{i=1}^{n} K_i(z_2^o)\eta_{1,i}}{\sum_{i=1}^{n} K_i(z_2^o)}. \tag{2.7}$$

With kernel smoothing methods and the two-step procedure explained here, estimates of both the parametric and nonparametric components of the model can be estimated. Research with kernel smoothing methods has suggested that the choice of bandwidth far outweighs the selection of the kernel function in estimation of conditional means (see Silverman (1986), Ullah (1988), or Härdle (1990) for details). We save the discussion of bandwidth selection until we discuss generalized kernel estimation of the fully nonparametric model.

Robinson (1988) noted that the additional information provided by the linear portion of the semiparametric model leads to $\sqrt{n}$-consistent estimation of the linear parameters (when correctly specified) and lessens the "curse of dimensionality". The benefit of a semiparametric specification over one that is nonparametric is that, if the hedonic price function is indeed partly linear, then estimates from this model are more efficient compared to a nonparametric specification. However, hedonic price theory rarely provides insight as to which attributes enter linearly (those with constant marginal valuations). Thus, the researcher is left with the crucial task of selecting which product characteristics will enter the linear portion of the model. Inclusion of variables in the linear component of the semiparametric model, when indeed they enter in a nonlinear fashion, may give biased results[5], especially if a second stage analysis is to be conducted, hence the validity of this specification needs to be tested.

All of the previous studies on estimation of hedonic price indices, while relaxing some aspect of the functional form of the hedonic price function, either estimate models with only continuous attributes[6] or allow the discrete attributes to enter into the functional form

[5]Not to mention that the marginal prices are constant instead of heterogeneous.

[6]This may induce an omitted variable bias if there are discrete attributes that are left out purposefully so that standard nonparametric methods may be used.

in a linear fashion. While Rosen's theory was predicated upon continuous attributes, his main point for discrete characteristics still holds, *because the bundle cannot be unpackaged and resold*, arbitrage possibilities are nonexistent, thus there is no reason to believe that the addition of the third bedroom, *ceteris paribus*, has the same marginal value as the addition of the sixth bedroom, *ceteris paribus*. However, until recently it was unclear how to incorporate both discrete and continuous variables into a nonparametric regression framework. Thus, this analysis has been missing from the hedonic price literature that attempts to relax functional forms. Here it turns out that even in what many would call a 'flexible' setting, the imposed linearity can still suffer from misspecification and lead to improper inferences about the underlying valuation structure of the housing market if the model is indeed entirely nonlinear.

2.2.2 Nonparametric Estimation

Although there are reasons to exclude shift variables from an unconstrained, unknown function in a semiparametric framework (see Pagan and Ullah 1999, pp. 198), we follow Rosen's (1974) suggestion and consider a fully nonlinear specification of the hedonic price equation. It is Rosen's belief in the nonlinearity of hedonic price models which provides the impetus for employing nonparametric estimation. Given the presence of categorical variables in many hedonic studies, we use recently developed nonparametric methods (Li and Racine (2004), Racine and Li (2004)) that smooth both ordered and unordered categorical data in nonparametric kernel regression. This approach is especially important here because we want to check the robustness/appropriateness of the findings from the semiparametric model.

To obtain estimates of the hedonic price function, as well as its derivatives, we employ Local-Linear-Least-Squares (LLLS). LLLS starts by approximating an unknown function with a first degree Taylor expansion and then employing standard kernel estimation techniques. This is demonstrated below. The unknown hedonic price function, assuming existence of a second derivative, is given as:

$$h(z) = h(z_o) + h'(z_o)(z - z_o) + \varepsilon_T, \tag{2.8}$$

where ε_T is the approximating error of the Taylor expansion. If this is coupled with our standard hedonic price model, Equation (3.1), we have

$$P = h(z_o) + h'(z_o)(z_i - z_o) + \varepsilon, \tag{2.9}$$

where $\varepsilon = \varepsilon^T + \varepsilon^h$. Similar to Equation (2.2), we can estimate our nonparametric model of the hedonic price function, Equation (2.9), using kernel methods, which are described in detail below. The benefit of specifying the model as a first order Taylor expansion is twofold. First, LLLS has a smaller bias than the LCLS estimator, see (Fan (1992, 1993)). Second, often in economic applications, the derivative of the function is of as much or more interest than the function itself. This is especially true with hedonic price estimation where knowledge of the slope of the function is paramount to a second stage analysis. LLLS estimation is best put in abstract terms for discussing estimation. Let $\beta_0(z_o) = h(z_o)$ and $\beta(z_o) = h'(z_o)$ refer to the vector of first derivatives or the hedonic price function evaluated at z_o. Then, kernel estimation of the hedonic price function amounts to minimization of

$$\sum_{i=1}^{n} \left(P_i - \beta_0(z_o) - \beta(z_o)'(z_i - z_o)\right)^2 K\left(\frac{z_i - z_o}{\omega}\right). \tag{2.10}$$

This is very similar to weighted least squares (WLS), with the weights being inversely proportional to each z_i's distance from z_o. If there were q_c continuous variables then the classical X matrix from OLS would contain a column of ones representing the intercept, $\beta_0(z_o)$, and q_c columns of *continuous* variables, $z_{i,j} - z_{o,j}$, where $j = 1, \ldots, q_c$ for the slope coefficients, $\beta_1(z_o), \ldots, \beta_{q_c}(z_o)$. The estimates of $\beta_0(z_o)$ and $\beta_1(z_o), \ldots, \beta_{q_c}(z_o)$ are

$$\begin{bmatrix} \hat{\beta}_0(z_o) \\ \hat{\beta}_1(z_o) \\ \vdots \\ \hat{\beta}_{q_c}(z_o) \end{bmatrix} = (Z'DZ)^{-1}(Z'DP) \tag{2.11}$$

where

$$Z = \begin{bmatrix} 1 & z_{1,1} - z_{o,1} & \cdots & z_{q,1} - z_{o,q_c} \\ 1 & z_{1,2} - z_{o,1} & \cdots & z_{q,2} - z_{o,q_c} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{1,n} - z_{o,1} & \cdots & z_{q,n} - z_{o,q_c} \end{bmatrix}, \qquad P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix},$$

and

$$D = \begin{bmatrix} K\left(\frac{z_{.,1}-z_o}{\omega}\right) & 0 & \cdots & 0 \\ 0 & K\left(\frac{z_{.,2}-z_o}{\omega}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K\left(\frac{z_{.,n}-z_o}{\omega}\right) \end{bmatrix}. \tag{2.12}$$

LLLS amounts to constructing kernel weights for each realization of a bundle of attributes with respect to the evaluation point and then performing WLS. As is common, this process is repeated over every data point, so, with n data points, the estimation is done n times, with new estimates of $\beta_0(z_o)$ and $\beta_1(z_o), \ldots, \beta_{q_c}(z_o)$ each time. The main difference between LLLS and WLS is that the weights are local (dependent upon the evaluation point) for LLLS and global for WLS (dependent upon the correction the weights are being used for). In this framework the estimated intercept represents the conditional mean of the unknown hedonic price function, while the slope coefficient estimates are the derivatives with respect to the continuous variables of the unknown hedonic price function. In the case where all of the attributes in the unknown function are continuous the product kernel used for LLLS is the same as in the semiparametric estimation, (Equation (2.5)).

However, rarely in economics does one encounter a dataset composed entirely of continuous variables. This is especially true for hedonic price analysis as more variables are often discrete than continuous. In the standard nonparametric framework with only continuous variables, the kernel function is designed to satisfy several properties. First, for large values of $(z_i - z_o)/\omega$, the kernel weight should be small; the farther away from the evaluation point the less weight an observation receives. Second, since the bandwidth ω goes to zero as the sample size grows towards infinity, the kernel weight goes to zero for all $z_i \neq z_o$. Finally, the kernel function must integrate to unity.

Much of this changes when the data are discrete. In order to smooth categorical data, a kernel function designed specifically for discrete data must be used. These types of kernels assign weights of unity when the discrete regressors z_i and z_o are identical, and weights which are functions of their associated bandwidth when they differ. Specifically, for variables that are discrete and display no natural ordering an appropriate kernel would be

$$k_i^{duo}(z_o) = \begin{cases} 1, & \text{if } z_i = z_o \\ \omega_{duo}, & \text{otherwise} \end{cases} \tag{2.13}$$

which is a variation of the discrete variable kernel proposed by Aitchison and Aitken (1976) for unordered categorical variables.[7] With variables that display a natural ordering, but are still discrete, the Aitchison and Aitken (AA) kernel is not appropriate to construct weights. Intuitively, the AA kernel only takes into account if two values of a variable are the same, treating the values as ordinal. If they are the same they get a weight of one, and if they are different the weight is given by the bandwidth. However, due to the ordinality of the AA kernel, nothing is conveyed about how different the values are, as does a continuous kernel, i.e. there is nothing reflecting the cardinal nature of the data. Given that ordered, discrete variables have a natural ordering, a kernel that recognizes that the data are discrete, but still takes into account "distance" is appropriate. Here, our kernel weighting function is given as

$$k_i^{do}(z_o) = \begin{cases} 1, & \text{if } z_i = z_o \\ \omega_{do}^{|z_i - z_o|}, & \text{otherwise} \end{cases} . \tag{2.14}$$

This kernel is also a variant of Aitchison and Aitken, but was used in almost this form by Wang and Van Ryzin (1981) (see their geometric weight function, pg. 302). The difference between the two kernels is that the kernel for the unordered data treats deviations from z_i equally, whereas the ordered data kernel weights the values differently depending on the distance between z_i and z_o. For instance, the number of bedrooms is an example of an ordered

[7]Note that for an unordered discrete variable the kernel weights do not sum to 1.

discrete variable, whereas a variable identifying separate locales within a metropolitan area would be representative of an unordered discrete variable.

Once again, with more than one variable, and more than one data type, the product kernel can be used. However, the formula given in (2.5) needs to be modified to account for varying kernel types. If there are q attributes, with q_{duo} of them being unordered discrete, another q_{do} of the ordered discrete variety, and the remaining q_c of them continuous attributes, our generalized product kernel is[8]

$$K_i\left(z_o\right) = \prod_{j=1}^{q_{duo}} k_i^{duo}(z_{o,j}) \prod_{j=q_{duo}+1}^{q_{do}+q_{duo}} k_i^{do}(z_{o,j}) \prod_{j=q_{do}+q_{duo}+1}^{q} k_i^{c}(z_{o,j}), \tag{2.15}$$

where $z_{o,j}$ is the j^{th} component of the attribute vector, $q = q_{duo} + q_{do} + q_c$ and $k_i^c(z_{o,j}) = k\left(\frac{z_{i,j}-z_{o,j}}{\omega_c}\right)$ is the Gaussian kernel described earlier.

After choosing the kernel function, the final issue to be resolved is the choice of bandwidths. As mentioned previously, it is believed that the choice of the continuous kernel function matters little in the estimation of the conditional mean, and selection of the bandwidths is considered to be the most salient factor when performing any type of semi- or nonparametric estimation. As indicated above, the bandwidths control the amount by which the data are smoothed. Large values of any or all of the bandwidths will lead to large amounts of smoothing, resulting in low variance, but high bias. Small values of the bandwidths, on the other hand, will lead to less smoothing, resulting in high variance, but low bias. This trade-off is a well known dilemma in applied nonparametric econometrics and thus we often resort to automatic determination procedures to estimate the bandwidths.

2.2.3 Bandwidth Selection

Although there exist many selection methods, one of the most appealing is Hurvich, Simonoff, and Tsai's (1998) Expected Kullback Leibler (AIC_c) criteria. This method chooses smoothing parameters using an improved version of a criterion based on the Akaike Information Criteria.

[8]Provided that the attribute vector has been arranged as discrete unordered, discrete ordered, and continuous variables.

AIC_c has been shown to perform well in small samples and avoids the tendency to under-smooth as often happens with other approaches such as Least-Squares Cross-Validation. Nevertheless, conventional methods, such as the AIC_c criteria, appear to perform well in finite samples. The basic idea behind the procedure is that we want to choose the bandwidths in order to achieve the best fitting model, while at the same time, being cautious not to interpolate the data. Specifically, we wish to minimize either

$$AIC_c^{np}\left(\omega_{duo},\omega_{do},\omega_c\right)=\log\left(\widehat{\sigma}^2\right)+\frac{1+tr(H)/n}{1-\left[tr(H)+2\right]/n} \tag{2.16}$$

for the generalized kernel bandwidths, or

$$AIC_c^{sp}\left(\omega\right)=\log\left(\widehat{\sigma}^2\right)+\frac{1+tr(H)/n}{1-\left[tr(H)+2\right]/n}, \tag{2.17}$$

for the semiparametric, continuous variable only bandwidths, where

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(P_j-\widehat{h}(z_i)\right)^2\\ &= \left(\frac{1}{n}\right)P'(I-H)'(I-H)P\end{aligned}$$

for the nonparametric model and

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(P_j-z_{1,i}'\hat{\beta}(\omega)-\widehat{h}(z_{2,i})\right)^2\\ &= \left(\frac{1}{n}\right)\eta_1'(I-H)'(I-H)\eta_1\end{aligned}$$

for the semiparametric model and I is an $n\times n$ identity matrix common across both methods. The H matrix is commonly called the smoother matrix as it is composed of all of the kernel weights for each data point across each evaluation point. In the nonparametric scenario we have, using data points as evaluation points

$$H=\begin{bmatrix}\frac{K_1(z_{1,1})}{\sum_{i=1}^{n}K_1(z_{1,j})} & \frac{K_1(z_{1,2})}{\sum_{i=1}^{n}K_1(z_{1,j})} & \cdots & \frac{K_1(z_{1,n})}{\sum_{i=1}^{n}K_1(z_{1,j})}\\ \frac{K_2(z_{2,1})}{\sum_{i=1}^{n}K_2(z_{2,j})} & \frac{K_2(z_{2,2})}{\sum_{i=1}^{n}K_2(z_{2,j})} & \cdots & \frac{K_2(z_{2,n})}{\sum_{i=1}^{n}K_2(z_{2,j})}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{K_n(z_{n,1})}{\sum_{i=1}^{n}K_n(z_{n,j})} & \frac{K_n(z_{n,2})}{\sum_{i=1}^{n}K_n(z_{n,j})} & \cdots & \frac{K_n(z_{n,n})}{\sum_{i=1}^{n}K_n(z_{n,j})}\end{bmatrix},$$

and similarly for the semiparametric model, the smoother matrix can be written as

$$H = \begin{bmatrix} \frac{K_1(z_{2,11})}{\sum_{i=1}^{n} K_1(z_{2,1j})} & \frac{K_1(z_{2,12})}{\sum_{i=1}^{n} K_1(z_{2,1j})} & \cdots & \frac{K_1(z_{2,1n})}{\sum_{i=1}^{n} K_1(z_{2,1j})} \\ \frac{K_2(z_{2,21})}{\sum_{i=1}^{n} K_2(z_{2,2j})} & \frac{K_2(z_{2,22})}{\sum_{i=1}^{n} K_2(z_{2,2j})} & \cdots & \frac{K_2(z_{2,2n})}{\sum_{i=1}^{n} K_2(z_{2,2j})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{K_n(z_{2,n1})}{\sum_{i=1}^{n} K_n(z_{2,nj})} & \frac{K_n(z_{2,n2})}{\sum_{i=1}^{n} K_n(z_{2,nj})} & \cdots & \frac{K_n(z_{2,nn})}{\sum_{i=1}^{n} K_n(z_{2,nj})} \end{bmatrix}.$$

Notice the slight difference between the H used for generalized kernel estimation, which includes all of the data in the kernel function, and the H used for semiparametric estimation, which only includes the continuous variables in the kernel function.

In this framework, as $\hat{\sigma}^2$ decreases, the fit of the model improves, but the second term, the "penalty for fit" term goes up, establishing the well known tradeoff between in-sample and out of sample prediction. The set of bandwidths that minimize the AIC_c function are those that are utilized in the final estimation. It should be noted that as the sample size grows and/or the number of regressors increases, the computation time of the AIC_c criteria increases exponentially. However, it is highly recommended that one use these types of techniques as opposed to a rule of thumb selection, especially with discrete data as no rule of thumb selection criteria exists. Another source of computational complexity is added by the number of bandwidths to be estimated. It is common that either a different bandwidth is specified for each variable or a separate bandwidth is specified for each data type. This is true whether one is using only continuous data or mixed discrete continuous data. The generalized kernel given in Equation (2.15) is written so that at most there are three bandwidths, one for discrete unordered, another for discrete ordered, and one for continuous data.[9] Also, for the semiparametric bandwidth(s) it is quite common that the same bandwidth(s) is used in the calculation of E_P, E_{z_1}, and E_{η_1}. As far as the author knows, there has been no research that

[9]Equation (2.15) can easily be generalized to accommodate a different bandwidth for every variable in the model by adding a subscript j to the bandwidths.

attempts to discern the benefits of estimating different bandwidths for the three conditional expectations.

Now that our discussion of semiparametric and nonparametric estimation is complete it is time to investigate the benefits reaped from nonparametric generalized kernel estimation as opposed to semiparametric estimation. We reconsider AG's highly cited semiparametric application to see how robust their results are when the linear attributes of the semiparametric model are allowed to enter in a nonlinear and unspecified manner.

2.3 An Empirical Comparison of the Two Methods

The data are composed of 546 observations from the Windsor, Canada housing market in 1987. At our disposal we have 11 covariates, bedrooms (6 groups), full bathrooms (4 groups), stories above ground (4 groups), garage places (4 groups), driveway (0/1), central air (0/1), recreational room (0/1), full finished basement (0/1), located in a preferred neighborhood (0/1), gas heated water (0/1), the lot size of the property (continuous), and a continuous regressand, the selling price of the house. To motivate the usefulness of the semiparametric specification, AG tested the appropriateness of their semiparametric model against a benchmark parametric model with a battery of specification tests. They rejected the linearity hypothesis in each of their tests, thereby advocating usefulness of their semiparametric specification. Before resorting to fully nonparametric estimation and a comparison of the results, it is useful to test whether a fully nonparametric estimation method is even needed. To determine the appropriateness of the specification employed by AG we use the partly linear specification test of Delgado and González Manteiga (2001, sec. 5.1). This test yields a bootstrapped p-value of 0.0675 which provides evidence against the semiparametric specification at the 10% level of significance, a reasonable level given our sample size and the methods we are employing. The bootstrapped p-value was calculated by estimating the semiparametric model using the bandwidth of AG.[10]

LLLS estimation results for the nonparametric model are provided in Table 2.1. Before discussion of the results we a mention of the construction of these marginal prices is in order. Due to the discrete nature of some of the attributes, they are not represented in the Taylor expansion used for LLLS. To determine the implicit value of any discrete attribute requires counterfactual analysis. So, if one wanted to determine the addition of a third bathroom, all the observations in the dataset with two bathrooms would require that the bathroom be changed from 2 to 3 and the predicted price increase would be noted as the implicit value of the third bathroom. This must be done one at a time, not simultaneously as counterfactuals require everything be held fixed except for the sole change of interest for a given observation.The table reports the mean coefficient with respect to each variable (along with its bootstrapped standard error in italics), as well as the coefficients at the 25th, 50th, and 75th percentiles (labeled Q_1, Q_2, and Q_3). The mean and quartile values of the coefficients on the discrete variables are noteworthy. The coefficients vary significantly over the quartiles aside from $BDMS = 2$ (perhaps due to the fact that there were only two houses in the sample with a single bedroom) and the insignificant $REC = 1$ (which was found to be significant in both the parametric and semiparametric procedures). This variation suggests that incorporating "intercept shifting" dummy variables may not be appropriate for discrete attributes. Similarly, the finding that a recreational room is not a significant attribute of a house's value should come as no surprise since a recreational room is a very subjective room of the house. Unlike a kitchen, bathroom, or deck, a recreational room is quite arbitrary as all it really signifies is another room in the house, which is captured through the lot size and bedrooms variables through the size of the house.

[10]This bandwidth differed slightly from the one calculated using the cross-validation procedure described previously. However, none of the resultant coefficient changes were significant so their bandwidth was used for all calculations relating to the semiparametric model.

Table 2.1: Generalized Kernel Estimation Results*

Variable	Mean	Q_1	Q_2	Q_3	Variable	Mean	Q_1	Q_2	Q_3
$DRV = 1$	0.051	0.025	0.043	0.075	$BDMS \geq 5$	0.075	0.036	0.058	0.107
	0.009	*0.002*	*0.009*	*0.011*		*0.013*	*0.008*	*0.018*	*0.012*
$REC = 1$	0.000	0.000	0.000	0.000	$FB = 2$	0.156	0.103	0.150	0.202
	0.003	*0.003*	*0.003*	*0.003*		*0.011*	*0.014*	*0.022*	*0.041*
$FFIN = 1$	0.113	0.076	0.155	0.286	$FB \geq 3$	0.294	0.231	0.307	0.361
	0.028	*0.012*	*0.028*	*0.036*		*0.029*	*0.029*	*0.029*	*0.025*
$GHW = 1$	0.186	0.075	0.155	0.286	$STY = 2$	0.061	0.030	0.055	0.084
	0.028	*0.012*	*0.012*	*0.035*		*0.004*	*0.003*	*0.004*	*0.004*
$CA = 1$	0.142	0.104	0.138	0.174	$STY = 3$	0.127	0.093	0.128	0.166
	0.021	*0.020*	*0.021*	*0.034*		*0.002*	*0.004*	*0.002*	*0.008*
$REG = 1$	0.078	0.055	0.081	0.109	$STY = 4$	0.197	0.167	0.185	0.250
	0.006	*0.001*	*0.013*	*0.013*		*0.012*	*0.008*	*0.009*	*0.008*
$BDMS = 2$	−0.014	−0.023	−0.016	−0.012	$GAR = 1$	0.026	0.005	0.024	0.046
	0.008	*0.011*	*0.008*	*0.008*		*0.002*	*0.003*	*0.003*	*0.003*
$BDMS = 3$	0.031	0.014	0.030	0.046	$GAR \geq 2$	0.029	0.011	0.030	0.041
	0.008	*0.007*	*0.008*	*0.008*		*0.003*	*0.002*	*0.003*	*0.003*
$BDMS = 4$	0.045	0.013	0.037	0.067	$\ln(LOT)$	0.404	0.320	0.390	0.473
	0.005	*0.007*	*0.008*	*0.007*		*0.077*	*0.069*	*0.078*	*0.069*

* The natural logarithm of price is used as the dependent variable in the regression. Q_1, Q_2, and Q_3 refer to the first, second, and third quartiles of the marginal price of the hedonic price function, respectively. The AIC_c criterion is used for bandwidth selection. Bootstrapped standard errors (199 replications) are listed in italics beneath each estimate.

Another benefit of the generalized kernel estimation procedure is that we can now analyze changes across ordered categorical variables without assuming a linear shift. For instance, the coefficient on $GAR = 1$ shows the counterfactual increase in the log price of a particular house when you add a one car garage to a house with no garage, ceteris paribus. Similarly, $GAR = 2$ shows the counterfactual increase in the log price of a particular house when you attach a two car garage to a house that previously has no garage, *ceteris paribus*. If the linear structure is appropriate, one would expect the coefficient on $GAR \geq 2$ (this is grouped because there are very few houses in the sample with a three car garage) to be at least twice that of $GAR = 1$. This is not the case. The mean coefficient goes from 0.026 when $GAR = 1$ to 0.029 when $GAR \geq 2$. That is, having a one car garage significantly increases the log price of a home, but the impact of an upgrade from a one-car to a two-car garage on the log price of a home is minimal. Finally, the coefficient on $\ln(LOT)$ is positive and significant at each quartile. This was not the case when we examined the marginal valuation of lot size from the AG semiparametric model[11], something not even reported in their paper. Each of these results suggests that the nonparametric procedure is more appropriate for this particular data set.

AG also claimed superiority of their semiparametric model by comparing goodness of fit criteria - the traditional R^2 vs. their pseudo R^2. However, this pseudo R^2 does not represent the proportion of the actual variation in the dependent variable explained by variation in the regressors, and therefore it is not correct to judge the appropriateness of the parametric fit by comparing these R^2 values. AG's measure of fit gives the impression that the semiparametric model fits the data vastly better than the linear model (0.923 vs. 0.684). However, this apparent great fit is illusory. If one were to calculate R^2 as the squared correlation coefficient between the actual and predicted dependent variable (as is done for the linear model), the fit of the semiparametric model is reduced to 0.738 using AG's bandwidth. Thus, the perceived fit of the semiparametric model turns out to be minimally better than the benchmark linear specification and the added flexibility afforded by the semiparametric specification does not coincide with a noticeably better fit. Incidentally, the fit of the nonparametric model is only slightly better than the semiparametric model, yielding a squared correlation coefficient of 0.756.

One common criticism by economists against the nonparametric method is that with several variables, *ceteris paribus* interpretations are usually not available. Given that economists are interested in policy analysis where only one variable is changed and then the impact of that change is analyzed, it would seem that nonparametric methods are not well adapted

[11]This was done using LLLS when estimating the unknown function.

to this type of research. However, nonparametric results can still be used to analyze economic situations of interest. One simply needs to look at subpopulations instead of the entire population simultaneously. Thus to determine the effect of an increase in square footage on housing prices, one could look over 3 bedroom houses, or houses with 3 bedrooms and 2 bathroom, or those houses with a garage, etc. Thus, selection of subpopulations of interest can provide a closer *ceteris paribus* interpretation with nonparametric methods rather than looking over the entire population.

To better understand the implicit price of a specific attribute, the estimated coefficients for each of the variables are used to examine different attribute scenarios for a house. For example, Table 2.2 shows that the value of an additional square foot of the lot ranges from \$4.73 to \$7.25 depending upon whether the house is a single story or a 4 story house, respectively. This change in the marginal value of the lot size suggests that lot size enters the model in a nonlinear fashion. Similarly, the implicit price of lot size ranges from \$4.74 to \$6.42 depending upon the number of bedrooms, from \$5.25 to \$7.05 for the number of full bathrooms, and from \$5.54 to \$6.21 depending upon the size or lack thereof a garage. The median values are also reported for each of these scenarios as averages are subject to outlier values. While the ranges for the median implicit price are smaller than their average counterparts, they still show variability, which suggests a nonlinear presence. Other interesting scenarios have been worked out and are reported in Tables 2.3 - 2.5.

The right hand column of Table 2.2 investigates the marginal price of lot size over the dichotomous individual attributes of our sample. Again, mean and median ranges for the estimated implicit prices are reported. For each attribute there appears to be an average increase in the implicit price of anywhere from \$0.36 (hot water heated by gas) to \$1.09 (having central air). The median differences are not always smaller than the average differences as they were in Table 2.2, \$1.23 for central air and \$0.55 for full finished basement, suggesting that the implicit value of another square foot of lot size depends heavily on the attributes of the house sitting on the land.

Table 2.2: Implicit Price Scenario 1: Looking Over All Houses

Features of House	Mean	Median	# Obs.	Features of House	Mean	Median	# Obs.
2 Bedrooms	4.74	4.61	136	Rec Room	6.19	5.61	97
3 Bedrooms	5.95	5.62	301	w/o Rec Room	5.61	5.10	449
4 Bedrooms	6.42	5.55	95				
				Full Fin Base	6.01	5.55	191
1 Story	4.88	4.73	227	w/o Full Fin Base	5.56	5.00	355
2 Story	5.96	5.47	238				
3 Story	6.32	5.86	40	Gas Hot Water	6.06	5.17	25
4 Story	8.32	7.52	41	w/o Gas Hot Water	5.70	5.25	521
1 Full Bath	5.25	4.85	402	Central Air	6.46	6.14	173
2 Full Baths	7.05	6.62	133	w/o Central Air	5.37	4.91	373
No Garage	5.54	5.13	300	Driveway	5.81	5.35	469
1 Car Garage	5.79	5.49	126	w/o Driveway	5.17	4.87	77
2 Car Garage	6.21	5.55	108				
				High Value Area	6.06	5.88	128
				Low Value Area	5.61	5.00	418

Table 2.3 provides the average and implicit price of lot size for the subpopulation of 3 bedroom houses in the sample, looking over garage spots, stories of the house and the presence of more than one full bath. Again, the average ranges are quite dispersed, as are the median ranges, suggesting the nonlinearity of lot size in the hedonic price function. We see a \$3.34 per square foot difference from a one story to a four story house, a \$1.68 increase in having additional full bathrooms, and a \$0.24 difference between having a garage versus a one car garage. The median ranges are similar and are not discussed further. The one theme across all three of the tables discussed is that the implicit price of lot size varies considerably depending upon the housing bundle.

The last scenario investigating the implicit price of lot size is the subpopulation of three bedroom/two story houses. Here we only look at how the implicit price varies across the

Table 2.3: Implicit Price Scenario 2: 3 Bedroom Houses

	Implicit Price of Lot Size		
Features of House	Mean	Median	# Obs.
1 Story	5.08	4.79	101
2 Story	6.06	5.88	143
3 Story	6.27	5.99	32
4 Story	8.42	8.06	25
1 Full Bath	5.53	5.03	225
> 1 Full Bath	7.19	7.28	76
No Garage	5.88	5.37	164
1 Car Garage	6.12	5.99	64
2 Car Garage	6.10	5.76	66

number of full bathrooms (given observation constraints). We notice that there is a $1.31 difference in the marginal value of lot size for having additional full bathrooms. Increasing a lot by 1000 square feet would lead to an additional increase in the price paid by $1,300 if the house had more than one full bathroom. For a home buyer putting this much extra into a 30 year mortgage would be considerable.

Table 2.4: Implicit Price Scenario 3: 3 Bedroom - 2 Story Houses

	Implicit Price of Lot Size		
Features of House	Mean	Median	# Obs.
1 Full Bath	5.77	5.47	111
> 1 Full Bath	7.08	6.91	32

The common theme represented in Tables 2.2 through 2.4 has been the implicit valuation of lot size, a continuous variable. However, previous studies have modeled continuous variables such as lot size in a nonparametric fashion and so these results could have been found in a semiparametric setting as well. What is of more interest is the implicit valuation of discrete variables in this generalized nonparametric framework. Table 2.5 shows the average value of adding an additional garage space to a home with a driveway and no garage. The addition of a one-car (two-car) garage, on average, is valued at \$2,601.62 (\$3,112.91) when compared to a similar house without a garage. This again shows that a linear specification for garage is inappropriate.

Table 2.5: Implicit Price Scenario 4: Houses with Driveways

	Implicit Price of A Garage Place		
Features of House	Mean	Median	# Obs.
1 Car Garage	2601.62	2087.63	108
2 Car Garage	3112.91	2195.97	107

Further pursuit of comparisons across models is shown in Table 2.6 for the 5 houses which have the median lot size of 4600 square feet. We compare the benchmark parametric model of AG, (their eq. 9, labeled as OLS(II) in the table), to the semiparametric and nonparametric models estimated in our paper (labeled as SP and NP, respectively). Each coefficient in the table is the implicit price of lot size $\partial P/\partial LOT = \widehat{\beta}_{\ln(LOT)} \cdot P/LOT$. For the first house (the least expensive one), the implicit price of lot size is nearly identical across all four models. This is not the case for the more expensive ones for which the implicit prices derived from the nonparametric model are the highest. For example, if the owner of the house that sold for \$127,000 purchased an additional 1,000 sq. ft. of land, it would lead to a considerable undervaluation of the land (the land value in the semiparametric and nonparmetric models would be \$11,146 and \$15,227, respectively – a difference of \$4,081). These differences for houses that sold for \$60,000 and \$75,500 are \$3,000 and \$2,817, respectively.

To determine whether the conclusions drawn from the nonparametric model are the result of over-fitting the data a small, out-of-sample fit exercise is performed by randomly removing n_1 = 46, 56, and 100 observations from the original dataset and constructing predictions

Table 2.6: Implicit Price of Median Lot Size

House	OLS(II)	SP	NP	Price	Price/Lot
1	2.926	2.887	2.820	43000	9.348
2	8.642	11.146	15.227	127000	27.609
3	3.402	5.085	4.437	50000	10.870
4	4.083	5.266	8.266	60000	13.043
5	5.137	6.042	8.859	75500	16.413

of those houses' selling prices using the remaining $n_2 = 500$, 490, and 446 observations to estimate the unknown function, respectively. We consider two different criterion, mean squared error ($MSE = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(P_i - \hat{P}_i\right)^2$) and mean absolute error ($MAE = \frac{1}{n_1}\sum_{i=1}^{n_1}\left|P_i - \hat{P}_i\right|$) and then repeated this process 199 times so as to avoid data mining. The results are presented in Tables 2.7 and 2.8.

Table 2.7: Out-of-Sample Fit: Mean Square Error[1]

n_1	Model	Average	Median	SD	IQR
46	SP	0.0503	0.0501	0.0112	0.0007
	NP	0.0530	0.0529	0.0115	0.0049
	NP/SP	1.0537	1.0559	1.0268	7.0000
56	SP	0.0498	0.0497	0.0102	0.0023
	NP	0.0526	0.0525	0.0106	0.0033
	NP/SP	1.0562	1.0563	1.0392	1.4348
100	SP	0.0505	0.0505	0.0074	0.0021
	NP	0.0530	0.0529	0.0077	0.0027
	NP/SP	1.0495	1.0475	1.0405	1.2964

We find that in terms of our criterion, the out-of sample fit of the full nonparametric method is almost equivalent to the predictive ability of the semiparametric model. The ratio of the MSE of the two models and the ratio of the MAE of the two models are essentially one.

The ratio of all criteria of the nonparametric model to the semiparametric model is decreasing as a larger sample is predicted upon.[12] Given the small sample size of the Windsor dataset a semiparametric specification may help circumvent the "curse of dimensionality," but for larger datasets with a higher likelihood of clustered observations it may prove quite erroneous to use such methods, and this is what we hope to put forth in this note, that while for this case the semiparametric may be a better fitting model, in terms of prediction, a generalized nonparametric approach should also be considered given the ability to smooth categorical data as well as the interpretive differences of the estimated marginal effects.

Table 2.8: Out-of-Sample Fit: Mean Absolute Error[1]

n_1	Model	Average	Median	SD	IQR
46	SP	0.1716	0.1737	0.0200	0.0041
	NP	0.1776	0.1771	0.0211	0.0086
	NP/SP	1.0350	1.0196	1.0550	2.0914
56	SP	0.1709	0.1703	0.0180	0.0051
	NP	0.1766	0.1758	0.0192	0.0073
	NP/SP	1.0334	1.0323	1.0667	1.4423
100	SP	0.1722	0.1724	0.0140	0.0060
	NP	0.1778	0.1775	0.0138	0.0049
	NP/SP	1.0325	1.0296	0.9857	0.8122

2.4 Concluding Remarks

Applied hedonic price theory is an important field in microeconometrics that continues to provide valuations of economic goods. Most applied hedonic papers have used rigid parametric specifications that severely limit their ability to properly value the attributes of the economic good in question. Attempts to use flexible parametric specifications are at best a shot in the dark because there exists no *a priori* theory that can tell how variables should

[12] One must not place too much trust in the MSE and MAE past several decimal places as they lose their accuracy.

enter in to the model, and which variables have constant marginal values. This point was illustrated using a well cited paper that uses the partly linear model. The goal of this chapter has been to show the attractiveness of nonparametric methods that have the ability to smooth categorical variables, in hopes of demonstrating the pitfalls of retreating to rigid model specifications in empirical analyses with no sound economic reasoning. For an example where economic theory proposes a model that stands up to nonparametric modeling see the investigation of the gravity model by Henderson and Millimet (2006).

A further extension to the research here would be the benefit of using a generalized nonparametric estimator in the first stage estimation coupled with the corresponding marginal prices being used in a second stage regression to uncover the structural demands. Here the relaxation of the functional form may avoid many of the identification pitfalls that other methods have succumbed to in the past. Also, the use of a nonparametric estimation procedure in the first stage analysis is consistent with EHN's argument that the hedonic price function is generically nonlinear and so generalized kernel estimation seems an appropriate econometric method for investigating hedonic indices. A more formal analysis of the estimation of the derivatives of the hedonic price index could be investigated along the lines of the studies by Cropper, Deck, and McConnell (1988) or Ohsfeldt and Smith (1985).

Chapter 3

Estimation of Hedonic Price Functions with Incomplete Information[1]

3.1 Introduction

In his seminal paper, Stigler (1961) attributed price variation to two overarching components: heterogeneity and ignorance (lack of information). As an example, consider the online textbook market. Different sellers sell the same text book on the internet at various prices. The product being sold is identical and yet different prices are observed. Thus, a buyer pays a higher price for a homogeneous product due to his/her ignorance. In the hedonic price literature, markets are sufficiently "thick" so as to eliminate price dispersion for identical goods. However, as a good becomes increasingly unique, markets become "thin" and the establishment of a single price for the product does not occur. Thus, prices may vary due to heterogeneity of goods and/or ignorance on the part of the buyer/seller. The uniqueness of goods becomes less of a reason for price dispersion as more characteristics are taken into account. As a result, ignorance becomes the prime factor causing price variations.

When purchasing goods, buyers face search costs in obtaining price information from different sellers. These search costs lead the buyer to search optimally in the sense that the marginal gain from another search is equal to its marginal cost (shadow price of search). If buyers knew of the the sellers with the lowest willingness to accept (WTA), as would be the case if the market was located at one central location and perfectly competitive, there would be no price variation except for the inherent randomness of the market and seller heterogeneity. This claim may be disputed by considering the study of Pratt, Wise, and Zeckhauser (1979) who find significant price dispersion, in terms of the coefficient of

[1] The basic ideas of this chapter are contained in Kumbhakar and Parmeter (2006).

variation, for 39 goods in what many would consider competitive markets. On the other hand, if sellers knew the buyers with the highest maximum willingness to pay (WTP), as is true for a first degree price discriminating monopolist, there would be no variation in prices except for noise in the data and heterogeneity among buyers.[2]

However, neither of the situations described above are common or even likely in today's marketplace. What is common is the existence of informational conduits through which both buyers and sellers can obtain information; the classified ads and the internet are just a few. While there is no way of knowing the highest WTP or the lowest WTA without some explicit advertisement of those prices, both buyers and sellers can gain information through search. However, given that search costs exist, market participants most likely will not become fully informed and price variations due to ignorance will exist even after controlling for product characteristics.

The extent of the price reductions obtained by buyers (from sellers) and the price increases received by sellers (from buyers) depends upon how well informed they are. Above, the lack of information has been labeled as ignorance. However, it may better this paper to use a milder terminology to characterize less than full information. Here, we call buyers and sellers who lack full information, deficient. As opposed to ignorance, deficient means that buyers and sellers may want to gather more information through search, but further search is costly and time consuming, and so the incomplete information that each possesses is sufficient to enter the market.[3]

The objective of this paper is two-fold. First, the hedonic price model is developed to incorporate incomplete information and shown to fit within the two-tier stochastic frontier (2TSF) framework(introduced by Polachek and Yoon (1987, 1996)), and a residual decomposition technique proposed by Kumbhakar and Parmeter (2005) is implemented to measure the impact of buyers' and sellers' information deficiency on observed market prices. The logic

[2]That is, buyers are not expected to have the same maximum WTP for the same product and so prices should vary across buyers.

[3]I thank Solomon Polachek for suggesting this taxonomy to me.

of the 2TSF is simple; the price of a good in any market is bounded above by the buyer with the highest WTP, the price at which that buyer is indifferent between paying for the good and not purchasing it, and bounded below by the seller with the lowest WTA, the price at which that seller is indifferent between selling the good and keeping it. Given the bounds on market price, it seems appropriate to employ the 2TSF technique to estimate a hedonic price function for a particular good. Second, we generalize the method to allow information deficiency to depend on buyer and seller characteristics. Empirically, we apply these techniques to estimate a hedonic price function for houses and examine the effects that characteristics such as being a first time buyer, being from out of town, having kids, and the like, have on housing prices. Since the impact of information deficiency on observed prices can be computed for each transaction, one can use these estimates to discern which types of buyers (sellers) are the most deficient in terms of paying (receiving) higher (lower) prices.

The remainder of the chapter is organized as follows. Section 3.2 briefly discusses the hedonic price model and demonstrates how incomplete information effects prices. In section 3.3 we show how to incorporate incomplete information into a hedonic model, which fits into the 2TSF model. Section 3.4 reviews the 2TSF model, focuses on how to construct measures of price efficiency (and the cost of information deficiency) for both buyers and sellers, and extends the model to allow for determinants of information deficiency. Section 3.5 presents the data and some results for the housing market. A summary of results and conclusions are given in section 3.6.

3.2 Incomplete Information and Hedonic Price Functions

Here we re-evaluate the hedonic price setup popularized by Rosen (1974). In these models, a good is composed of a set of characteristics. The implicit values of these characteristics constitute the market value of the good. Each of these characteristics has a shadow price and the price of the good is the affine transformation of the shadow prices and the requisite characteristics. In what follows we consider the Rosen (1974) model.[4]

Bid functions are constructed to discern the price a buyer would pay for a good with certain attributes for a fixed level of utility and income. It indicates the highest price a buyer is willing to pay for the good. Utility is maximized when the bid function is equal and tangent to the market price function, the minimum price a buyer must pay in the market. Similarly, firms have offer functions for the good in question, given a certain level of profit. The offer functions indicate the lowest price they will accept for the product in question. If the market is competitive, sales take place where an individual seller's offer function 'kisses' an individual buyer's bid function. In other words, the locus of the points of tangency between the offer functions $\psi(z, \pi)$ (for different levels of profit, π) and bid functions $\theta(z, u)$ (for different levels of utility, u) traces the hedonic price function $P(z)$, of the good. This is shown in Figure 3.1.

However, this model implicitly assumes that there is full information in the market and so buyers and sellers are perfectly aligned in the market, resulting in tangency between bid and offer curves. Ideally, the hedonic price function represents a Pareto optimum because buyers are paying the lowest possible price in the market for their given bundle selection, while sellers are receiving the highest price in the market for their selected bundle, and neither benefits from switching to another transaction . In this setup there would be no rationale for price variation except that which is due to inherent heterogeneity of the good.[5] With full information, every buyer knows of the seller with the lowest WTA and every seller knows of the buyer with the highest WTP. This forces the market to generate a unique price for each bundle, z, in the market.

However, with incomplete information the market price is affected by the levels of information that buyers and sellers possess specifically because there is no rationale for the buyer

[4]Given the detailed discussion of the standard hedonic model of Rosen covered in Chapter 1, the description here will be brief.

[5]This would be the error term from the regression of prices on the characteristics. Measurement error may also arise, but this is an econometric issue, not a theoretical one.

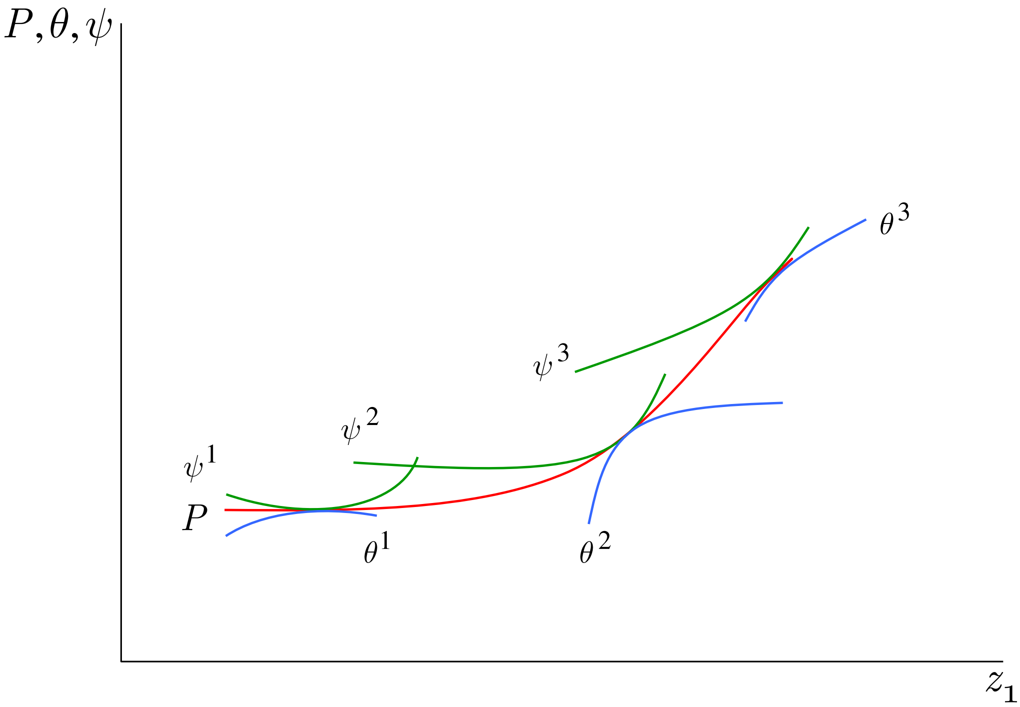

Figure 3.1: Market Equilibrium in Rosen's Framework

with the highest WTP to purchase from the seller with the lowest WTA. That is, for a given z, there is a gap between the highest WTP and the lowest WTA, and this gap is a source of price variations that are observed in the market. To demonstrate this lets consider each side of the market individually, and then combine them to make the intuition easier to handle. We begin with the buyers' side of the market shown in Figure 3.2.

Here we have a hedonic price function for buyers, $P_b(z_1)$, that is defined as the upper envelope for the bid functions θ^1, θ^2, and θ^3, similar to Rosen's framework. However, there are other buyers in the market that are looking to purchase the good at lower bids, θ^4, θ^5, and θ^6. In a world of perfect competition and full information these lower bids are not relevant to sellers as they know of bids θ^1, θ^2, and θ^3. It is this upper envelope that sellers with full

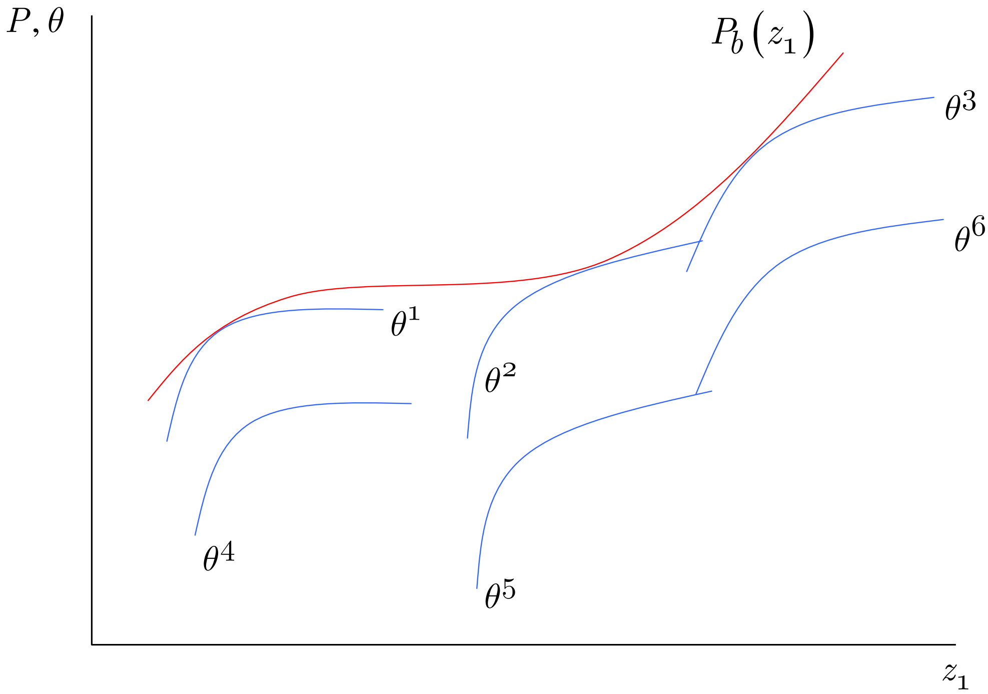

Figure 3.2: Sellers' Viewpoint of the Buyers' Side of the Market with Incomplete Information

information would align themselves with. The situation on the sellers' side of the market is similar and is depicted in Figure 3.3.

Here we have a hedonic price function for sellers, $P_s(z_1)$, that is defined as the lower envelope for the offer functions θ^1, θ^2, and θ^3, and is, again, similar to Rosen's framework. However, there are other sellers in the market that are looking to provide the good at lower offers, θ^4, θ^5, and θ^6. In a world of perfect competition and full information these lower offers are not relevant to buyers as they know of offers θ^1, θ^2, and θ^3. In Rosen's model we would have $P_b(z_1) = P_s(z_1)$ and price fluctuations would only exist due to heterogeneity. However, with incomplete information, $P_b(z_1) \neq P_s(z_1)$ and a more detailed analysis is required to understand what is happening in the differentiated goods market.

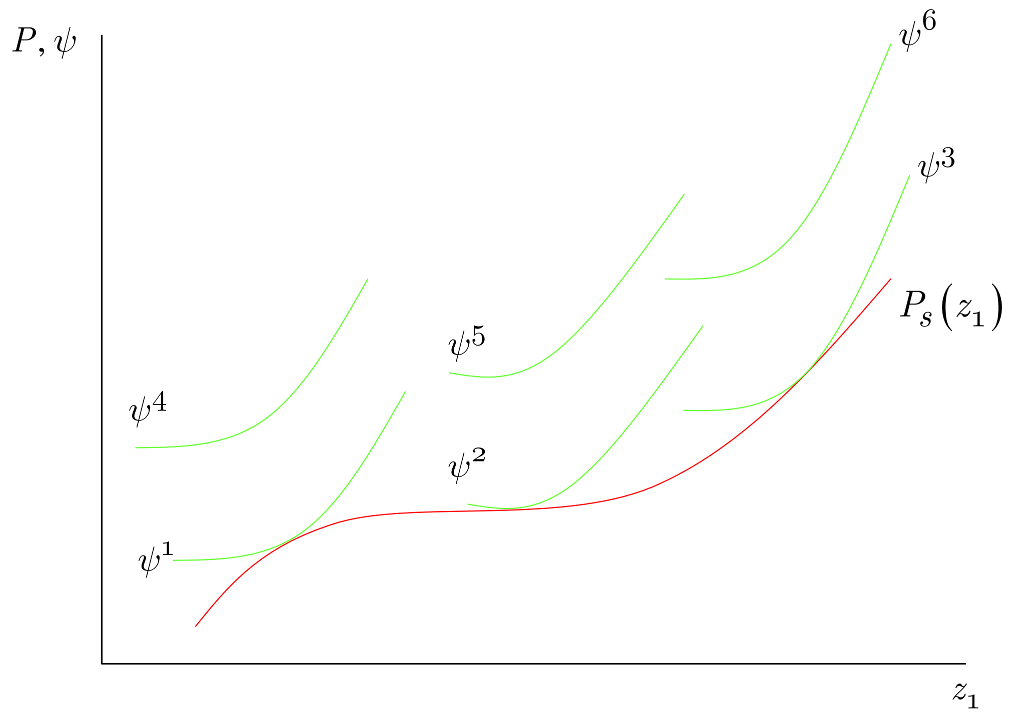

Figure 3.3: Buyers' Viewpoint of the Sellers' Side of the Market with Incomplete Information

In Figure 3.4 we place Figures 3.2 and 3.3 on top of one another. Somewhere in the space on or between $P_b(z_1)$ and $P_s(z_1)$ buyers and sellers will align and transactions will take place. Where these transactions occur will depend upon which agent (buyer or seller) has more relative information. These transactions will result in an equilibrium in two senses. One is that the good will be exchanged, the classical equilibrium, but the other is that both the buyer's bid function and the seller's offer curve will be tangent at the transaction. This is due to the fact that both buyers and sellers search optimally i.e. the marginal benefit of search is less than the marginal cost of search. If one were to dispute this then transactions would still occur, but both agents may be in a position where further search would be beneficial.

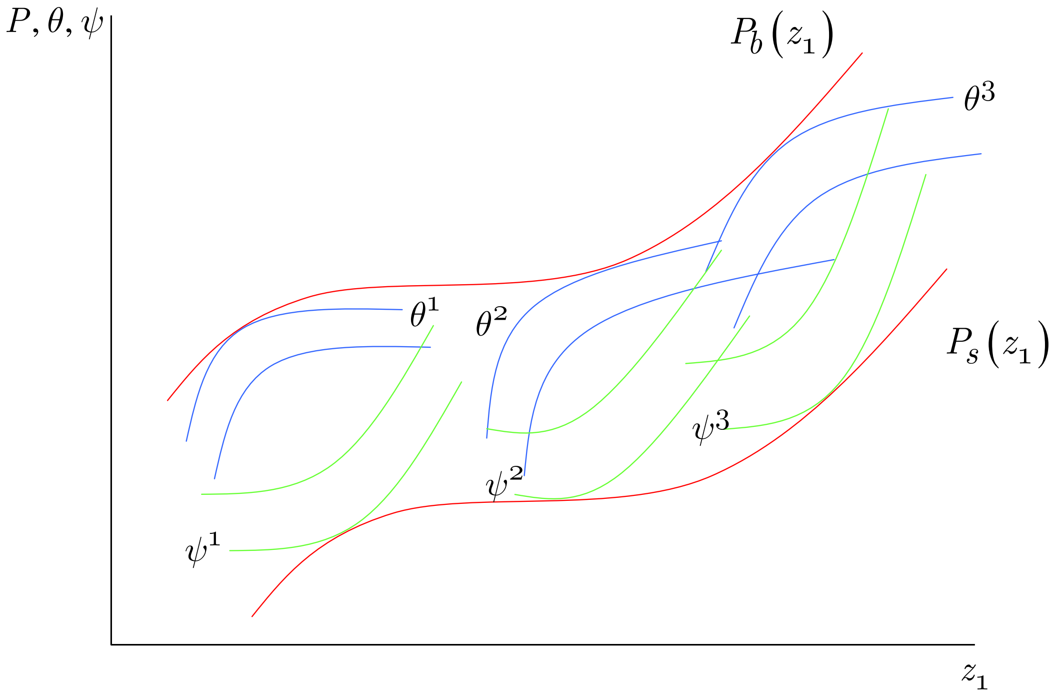

Figure 3.4: Structure of a Differentiated Good Market When Incomplete Information Exists

The equilibrium (with incomplete information) hedonic price function is bounded between these two full information envelopes, and is denoted by $P_h(z_1)$ in Figure 3.5.

It can be seen from this representation that the equilibrium hedonic price functions does not have to be equidistant from the bounds throughout its range. For equilibrium transactions of z_1 we see that at z_1' the seller is closer to the buyer envelope, while for z_1'' both agents are equidistant from each others corresponding envelopes, while for z_1''' the buyer is closer to the seller envelope. How far apart the two envelopes are and the relative position of the equilibrium hedonic price function is from both of them will depend upon myriad factors including but not limited to the product in question, the setup of the market, the characteristics of the buyers and the characteristics of the sellers. In the incomplete

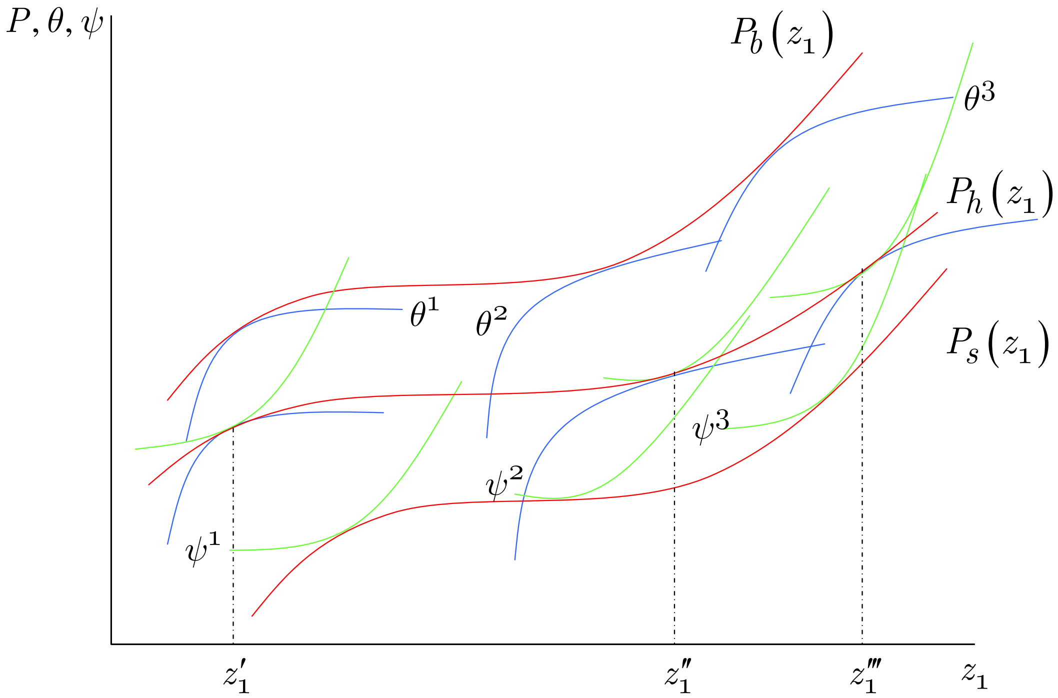

Figure 3.5: Equilibrium in a Differentiated Good Market When Incomplete Information Exists

information framework the observed price will depend on the attributes of the product, but also the information possessed by both buyer and seller for a given transaction.

What do these price functions, $P_b(z)$ and $P_s(z)$, imply about the market equilibrium price that is observed? The $P_b(z)$ function indicates that equilibrium prices cannot be higher than this function, otherwise no sale will result. Sellers wish to charge prices corresponding to this upper envelope. However, incomplete information bars sellers from doing so. Additional information about potential buyers can be obtained through market research and further interactions, which may lead to an increase in the prices received. The same story holds on the buyers' side. Here buyers attempt to find the "lowest prices" available in the market, i.e., prices corresponding to points on $P_s(z)$. The presence of search costs limits their ability to find out what the lowest price of the product is in the market. Buyers get additional information about market prices by soliciting prices from more sellers.

If one were to simply consider the hedonic price function without taking account of the highest and lowest prices just discussed, nothing can be said about the impact of the level of information that any particular buyer or seller possesses on the observed price. Thus, an estimation strategy that incorporates information deficiency in the market would seem appropriate. In the following section we show how the 2TSF approach can allow one to take advantage of the fact that incomplete information exists in the market as well as analyzing the effect that it has on the observed price relative to the attribute only price that could have been paid.

3.3 Hedonic Models with Incomplete Information

Consider the standard hedonic price equation that is typically specified as

$$P_h = h\,(z) + v, \tag{3.1}$$

where p_h is the hedonic price (or the logarithm of the hedonic price), z is a vector of product characteristics influencing the overall value of the good in the market, and v represents random noise and measurement error in price. The hedonic price function in (3.1) corresponds to the full information model developed by Rosen (1974). Nowhere in this setup are the buyers with the highest WTPs or the sellers with the lowest WTAs accounted for.

If we recast the hedonic price model in terms of these price bounds then from a seller's standpoint the market price (P_s) can be represented as

$$P_s = \bar{P} - u, \tag{3.2}$$

where $\bar{P}$ represents the highest WTP in the market and $u \geq 0$ can be viewed as the tax imposed on the seller for having incomplete information. From a buyer's point of view the market price (P_b) received for the good can be written as

$$P_b = \underline{P} + w, \tag{3.3}$$

where $\underline{P}$ represents the lowest WTA and $w \geq 0$ represents the tax paid by the buyer due to information deficiency. Thus, information deficiencies can be viewed as taxes paid by both buyers and sellers.

Given that market prices are bounded from above and below, we need a technique that takes these bounds into account and provide estimates concerning the effects of buyer and seller information deficiencies. In the 2TSF approach, the effects of incomplete information are captured because the market price is not identical to the hedonic price of the good. The observed market price must be adjusted for the incomplete information that is present at the time of the transaction. However, given that the levels of information possessed by buyers and sellers are unknown, an alternative strategy is needed in order to correctly interpret the problem at hand. We incorporate the effects of incomplete information into the standard hedonic price framework as follows.

For a market transaction to occur we require that the market price, P_m equals both the price paid by a buyer and the price received by a seller, $P_b = P_s \equiv P_m$ which yields the following equality

$$\underline{P} + w = P_m = \bar{P} - u. \tag{3.4}$$

We exchange u and w across the equality which gives,

$$\underline{P} + u = P_m + u - w = \bar{P} - w. \tag{3.5}$$

The left hand side can be viewed as the adjusted hedonic price of the seller while the right hand side can be viewed as the adjusted hedonic price of the buyer. These adjusted hedonic prices account for information that is possessed by market participants at the time of an exchange. Given that u, w, $\bar{P}$, and $\underline{P}$ are all unobserved, estimating either the LHS or the RHS of (3.5) would provide quite a challenge.

To obtain a regression equation from (3.5), we set the hedonic price of the good, $h(z)+v$, equal to the middle term of (3.5). Thus, we can express prices observed in the market as:

$$h(z) + v = P_m + u - w \Rightarrow P_m = h(z) + v + w - u. \tag{3.6}$$

Equation (3.6) shows that the market price of a good is constituted of (i) the implied value of the characteristics $h(z)$, (ii) inherent heterogeneity and random noise (v), and (iii) the costs of incomplete information to the buyers (w), and sellers (u). With this representation one can find not only marginal values of attributes, but the impact of incomplete information on prices (labeled as the tax or the cost of incomplete information). The market price should be the same as the hedonic price (aside from v) when (i) there is no information deficiency on the part of either buyers or sellers, or, (ii) the cost of buyers incomplete information completely offsets the cost of sellers incomplete information (i.e., $w - u = 0$). The standard hedonic price models take the market price as the true valuation of the good. In doing so the asymmetry in the costs of the incomplete information possessed by the buyers and sellers is discarded and consequently the only explanation of price dispersion is heterogeneity.[6] Our model incorporates information deficiency from both sides of the market, as suggested by Rothschild (1973), which should help characterize the nature of price dispersion better than the standard model that assumes away information deficiency.

Collecting the three separate error components in (3.6), we write the estimating equation as,

$$P_m = h(z) + v + w - u = h(z) + \varepsilon. \tag{3.7}$$

Here $\varepsilon = v + w - u$ is a three component composite error term. One can view u and w as the incomplete information tax (cost of incomplete information) paid by the seller and buyer, respectively.[7]

[6] Indeed, as pointed out by Harding, Rosenthal, and Sirmans (2003), the fact that housing prices display seasonal variation suggests a weakness of the standard hedonic price model for the housing market.

[7] Incomplete information tax and the cost of incomplete information are interchangeably used.

The parameters of $h(z)$ in (3.7) can be obtained from a simple regression once a parametric form of $h(z)$ is chosen. Since u and w are one-sided, $E(\varepsilon)$ may not be zero, even if $E(v) = 0$. Consequently, the OLS estimate of the intercept will be biased.[8] However, the OLS procedure will give unbiased and consistent estimates of the slope coefficients so long as the error components are homoscedastic and uncorrelated with any of the good's characteristics. If the ultimate goal is to determine the marginal valuation of particular characteristics then estimation of standard hedonic price models are sufficient for these purposes. However, if the objective is to learn about the effects of incomplete information on price variations, then a model such as (3.7) is more appropriate.

Before moving on it is worth noting the difference of the framework proposed here as opposed to some of the previous attempts to model information and its impact on hedonic prices. Past research has attempted to model the impact of knowledge by using dummy variables to capture the amount of supposed information that buyers/or sellers have, and then use the signs of the coefficients corresponding to these dummy variables to determine the impact that information has on the overall product price.[9]

For example, Sirmans and Turnbull (1993) use dummy variables for 'first time' house buyers and 'out of town' house buyers to test the hypothesis that these groups are less informed than 'in town' and 'repeat' house buyers. However, the hedonic price index represents a simultaneity between buyers and sellers. But none exists for these variables; sellers are not offering the 'out of town' attribute to house buyers. This makes the interpretation of the coefficient of the information dummy variables difficult if not impossible, and it does not formally show how information is being captured within the hedonic price framework. The 2TSF framework will provide a way to assess the impact that each agents information has on the price paid as well as to determine a relative measure of incomplete information, which is consider in the following.

[8] Note that although $E(u)$ and $E(w)$ are non-zero, $E(w-u)$ might be zero. If this happens then the OLS estimator of the intercept will also be unbiased.

[9] Not all studies have taken this route. Others have used truncation models or have had agents maximize in the presence of uncertainty, such as maximizing *expected* utility as opposed to utility.

3.4 The Two-Tier Frontier Method

In this section the 2TSF framework developed by Polachek and Yoon (1987) is reviewed and derive the conditional expectations for the information deficiency effects using Jondrow et. al.'s (1982) methodology. These conditional expectations will be the base for determining the impact of incomplete information as well as the relative price increase due to differing amounts of incomplete information across buyer and seller. A test for no impact of incomplete information is also discussed.

3.4.1 The Log Likelihood Function

Standard (single-tier) stochastic frontier models include either u (for production, revenue, and profit functions) or w (for a cost function) depending on the assumed behavior of firms. In the present case, we have two frontiers, viz., an outer frontier $\bar{P}_i = P_{m,i} + w_i$, which represents the highest price that a buyer (in the i^{th} transaction) is willing to pay, and an inner frontier $\underline{P}_i = P_{m,i} - u_i$, which represents the lowest price that the seller (in the i^{th} transaction) is willing to accept. These two frontiers are imbedded in (3.7).

To estimate (3.7) we use the maximum likelihood (ML) method. The ML method we propose is based on the following distributional assumptions of the error components, viz., v, u and w: (i) $v_i \sim i.i.d.\ N\left(0, \sigma_v^2\right)$, (ii) $u_i \sim i.i.d.\ Exp\left(\sigma_u,\, \sigma_u^2\right)$[10] (iii) $w_i \sim i.i.d.\ Exp\left(\sigma_w,\, \sigma_w^2\right)$, along with the assumption that each of these error components is distributed independently of one another and from each of the regressors. The exponential distributions for u and w capture the fact that the probability of low costs of incomplete information for buyers and sellers are high (the area near zero values of u and w being the highest), which one would expect if markets work well.

Based on the above distributional assumptions, it is straightforward (but tedious) to derive the pdf of ε_i, $f\left(\varepsilon_i\right)$, which is

[10] Here $Exp\left(\sigma_z,\, \sigma_z^2\right)$ denotes a random variable z that is exponentially distributed with mean σ_z and variance σ_z^2.

$$f\left(\varepsilon_i\right)=\frac{\exp\left\{\alpha_i\right\}}{\sigma_u+\sigma_w}\Phi\left(\beta_i\right)+\frac{\exp\left\{a_i\right\}}{\sigma_u+\sigma_w}\Phi\left(b_i\right) \quad (3.8)$$

where

$a_i = \frac{\sigma_v^2}{2\sigma_w^2}-\frac{\varepsilon_i}{\sigma_w}$, $b_i = \frac{\varepsilon_i}{\sigma_v}-\frac{\sigma_v}{\sigma_w}$, $\alpha_i = \frac{\varepsilon_i}{\sigma_u}+\frac{\sigma_v^2}{2\sigma_u^2}$, $\beta_i = -\left(\frac{\varepsilon_i}{\sigma_v}+\frac{\sigma_v}{\sigma_u}\right)$ and $\Phi\left(.\right)$ is the cumulative distribution function of the standard normal variable. The likelihood function for a sample of n observations is the product of the $f(\varepsilon_i)$ in (3.8), $\prod_{i=1}^{n} f\left(\varepsilon_i\right)$.

The log likelihood function for a sample of n observations identically distributed according to (3.8) is given as

$$\ln L\left(x;\theta\right)=-n\ln\left(\sigma_u+\sigma_w\right)+\sum\nolimits_{i=1}^{n}\ln\left[e^{\alpha_i}\Phi\left(\beta_i\right)+e^{a_i}\Phi\left(b_i\right)\right], \quad (3.9)$$

where $\theta=\{\delta,\sigma_v,\sigma_u,\sigma_w\}$ and δ represents the parameters associated with the hedonic function $h(z)$. The estimates of all the parameters can be obtained by maximizing the log likelihood function. It should be noted that identification of σ_u and σ_w is achieved due to the fact that they appear in the likelihood equation in distinct parts. We use the exponential-exponential-normal set up here for tractability, although one could use distributions such as truncated normal or half normal, as is done in the traditional single-tier frontier models.

3.4.2 Buyer's and Seller's Costs of Incomplete Information

The main goal of estimating a stochastic frontier function is to obtain estimates of observation-specific inefficiency. In the present case we wish to determine by how much more a buyer pays and how much less a seller receives for having incomplete information.[11] To explain these concepts we reconsider equations (3.2) and (3.3), and note that a buyer's price efficiency is the ratio of the lowest WTA to the observed price. That is,

$$PE_{Buyer_i}=WTA_i/\text{observed price}_i=\exp\left\{-w_i\right\}, \quad (3.10)$$

[11] Note that these are all potential in the sense that they are based on WTP and WTA.

while a seller's price efficiency is the ratio of the observed price to the highest WTP and is represented by

$$PE_{Seller_i} = \text{observed price}_i / WTP_i = \exp\left\{-u_i\right\},. \tag{3.11}$$

Following the single-tier frontier approach (Jondrow et al (1982)), we estimate (3.10) and (3.11) using their conditional means, viz., $E\left(e^{-u_i} \mid \varepsilon_i\right)$ and $E\left(e^{-w_i} \mid \varepsilon_i\right)$. These formulae are:

$$E\left(e^{-u_i} \mid \varepsilon_i\right) = \frac{\lambda}{1+\lambda}\frac{1}{\chi_{2i}}\left[\Phi\left(b_i\right) + \exp\left\{\alpha_i - a_i\right\}\exp\left\{\sigma_v^2/2 - \sigma_v\beta_i\right\}\Phi\left(\beta_i - \sigma_v\right)\right] \tag{3.12}$$

and

$$E\left(e^{-w_i} \mid \varepsilon_i\right) = \frac{\lambda}{1+\lambda}\frac{1}{\chi_{1i}}\left[\Phi\left(\beta_i\right) + \exp\left\{a_i - \alpha_i\right\}\exp\left\{b_i\sigma_v - \sigma_v^2/2\right\}\Phi\left(b_i + \sigma_v\right)\right]. \tag{3.13}$$

To estimate the cost of information deficiency of a particular buyer or seller, we compute $1 - PE_{Buyer_i} = 1 - e^{-w_i}$ and $1 - PE_{Seller_i} = 1 - e^{-u_i}$, using the above formulae for e^{-w_i} and e^{-u_i}. Note that if a buyer (seller) were fully informed he/she would find the lowest price regardless of sellers' (buyers') incomplete information. When our dependent variable is in logarithms, one can interpret w (u) (when multiplied by 100) as the percentage increase (decrease) in price that a buyer (seller) pays (receives), especially when w (u) is small.

3.4.3 Heterogeneity in Buyer's and Seller's Deficiency

So far we have assumed that the distribution of w (u) is identical for all buyers (sellers). Since the costs associated with incomplete information are likely to differ across buyers and sellers, the 2TSF model has to be extended to incorporate the possibility of systematic differences in w and u. In view of this, our objective in this subsection is to allow the parameters of the distribution of w (u) to depend on buyers' (sellers') characteristics, so that incomplete information taxes can differ systematically. For example, if it is believed that first time home buyers have less information than repeat home buyers (Turnbull and Sirmans 1993), one incorporates this information directly into the hedonic price function by adding a dummy variable and performs a statistical test. In our framework, we can go further and assert that the distribution of w is different for repeat buyers and first time buyers. This can be done by allowing the mean (standard deviation) of w (u) to depend on buyers' (sellers') characteristics.

To allow for the possibility that a vector of exogenous variables can influence the cost of incomplete information, we allow the means of w and u to be functions of buyers' and sellers' characteristics (z_w and z_u variables), respectively. Thus, we specify σ_w and σ_u as

$$\sigma_w = e^{\delta'_w z_w} \quad \text{and} \quad \sigma_u = e^{\delta'_u z_u} \tag{3.14}$$

where z_w and z_u are vector of buyers' and sellers' characteristics (including intercepts).[12] If we believe that a particular characteristic lowers (increases) the expected cost of incomplete information, such as being a repeat buyer, then we would expect the associated coefficient to be negative (positive). The standard likelihood ratio (LR) test can be used to test whether some of these characteristics influence σ_w and σ_u. Thus, the benchmark model becomes a special case of this extended model, when the coefficients of z_w and z_u are jointly zero (except the intercepts). The log likelihood function of this extended model is same as the benchmark model. The only difference is that σ_u and σ_w in (3.9) are to be replaced by the functions in (3.14). Similarly, to estimate e^{-w_i} and e^{-u_i} in the extended model, we use the formulae in (3.12) and (3.13) but replace σ_u and σ_w by the functions given in (3.14).

3.4.4 Testing for Information Deficiency

While the estimation of mean and individual effects of incomplete information are useful, it is also interesting to examine whether these effects exist at all. From estimation of Equation (3.9), simple Wald statistics can be constructed to determine the individual significance of estimates of σ_u and σ_w. However, a broader investigation of incomplete information requires joint consideration of the magnitudes of σ_u and σ_w.

[12]This formulation is similar to the scaling method of Wang and Schmidt (2002) in a single-tier frontier model.

Unfortunately, testing for joint significance is not as simple as constructing a likelihood ration test as the null hypothesis of $\sigma_u = 0$, and $\sigma_w = 0$, lies on the boundary of the parameter space. Testing constraints that lie on the boundary of the parameter space require the use of mixed χ^2 distributions as opposed to the standard textbook treatment of likelihood ratio statistics. Much work has been done on testing inequality constraints and boundary constraints with likelihood ratio tests, see Chernoff(1954), Moran(1971), Chant (1974), Gourieroux, Holly, and Montfort (1982), and Self and Liang (1987) for purely statistical insights. Econometric applications relating to error components models can be found in Miller (1977), Honda (1985), Baltagi, Chang and Li, (1992), and Coelli (1995).

Typically an r-restriction hypothesis test using a likelihood ratio statistic has an asymptotic χ^2 distribution with r degrees of freedom. When boundary and inequality constraints are present the asymptotic distribution is mixed χ^2 with mixing components dependent upon the angles between the alternative subspaces that the parameters lie within. With r restrictions that are all boundary or inequality constrained, the appropriate asymptotic distribution is,

$$\chi_A^2 = \sum_{i=0}^{r} w(\theta, i)\chi^2(i), \tag{3.15}$$

where θ is the angle between appropriate subspaces in the alternative hypothesis, $\chi^2(i)$ is a χ^2 distribution with i degrees of freedom, and $\chi^2(0)$ is the point mass at zero. When the restrictions on the corresponding parameters are independent of one another (i.e. the alternative subspaces are orthogonal) the weights are $w(\theta, i) = \begin{pmatrix} r \\ i \end{pmatrix} /2^r$, see Gourieroux et. al. (1982, pg. 79)and Self and Liang (1987, pg. 609).

For our purposes we will be testing that the effects of incomplete information are zero for both buyers and sellers. The independence of σ_u and σ_w means that our weights for Equation (3.15), become $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively for the two restrictions. This yields an asymptotic distribution for our likelihood ratio test statistic of $\frac{1}{4} \cdot \chi^2(0) + \frac{1}{2} \cdot \chi^2(1) + \frac{1}{4} \cdot \chi^2(2)$. The critical values for this distribution at the 1, 5, and 10% significance levels are 7.289, 4.321, and 2.952, respectively (see Baltagi et. al. (1995, pg. 99)). It is worth noting that because of the form of the likelihood function in Equation (3.9), one cannot simply set σ_u and σ_w to zero explicitly as the likelihood function would be undefined. Rather, given the assumption of normality, testing for no effects of incomplete information is equivalent to estimation using standard maximum likelihood with an assumed normal distribution.

3.5 A Housing Market Application

We apply the 2TSF techniques developed in the preceding sections to examine the cost of incomplete information on housing prices. Our data[13] for this study comes from the American Housing Survey (AHS) which was recently investigated by Harding, Rosenthal, and Sirmans (2003), (HRS hereafter).

3.5.1 Data

The data for the AHS is collected every two years and contains not only the traditional characteristics of the house, such as the number of rooms and square footage, but characteristics of the primary homeowner. Thus, attributes of buyers or sellers that may be believed, *a priori*, to influence the extent of search that takes place, such as having kids, wealth, not being from the local area, being a first time buyer, and the like, can be incorporated into the analysis in a parsimonious manner that will allow us to investigate the impact of search costs on housing prices.

We use a mix of house attributes and local area characteristics in the hedonic price function. The attributes used to explain housing prices are: square footage of the floor space, an indicator of whether the floor space variable was top coded, the total number of rooms in the house, the total number of bathrooms in the house, dummy variables for the age of the house, a quality control value that indicates if the house is deemed inadequate prior to being sold and dummy variables that capture the effect of the house being either a single family attached or a single family detached house. The local area characteristics are: dummy

[13] Stuart Rosenthal kindly provide the data set used in this application.

variables for the area the house is located in (city center, urban suburban, rural, other urban area), dummy variables for the population of the MSA the house is located in (greater than 7 million people, between 3 and 7 million, between 1 and 3 million, and not in a MSA), and a climate control variable that ranges from 1 to 6. We also included time dummys to capture any year effects that may be effecting the housing market. For a detailed description of the data we refer the readers to HRS.

We use this set of variables in the benchmark model to estimate the hedonic price and then determine the cost of incomplete information for buyers and sellers. We also consider an extended model in which characteristics such as race, marital status, gender, having children, having a college education, owning a business, age, income, being a first time buyer, and being a buyer from out of town are used as correlates in the hedonic function. Furthermore, we allow these characteristics to affect prices indirectly through the means of u and w. In other words, information taxes are allowed to vary systematically across buyers and sellers.

3.5.2 Results from the Benchmark Model

Estimates of a linear hedonic price function from the benchmark model and the parameters associated with the distributions of v, u, and w are presented in Table 3.1. The coefficient estimates are comparable to those in HRS (Table A2 column 2), which is to be expected since the OLS estimates of the slope coefficients are unbiased. Our main interest is the estimates of the parameters, σ_u and σ_w, which are individually statistically significant, suggesting that both buyers and sellers in the housing market are incompletely informed. A joint test of significance yields a likelihood ratio of $-2 \cdot [-4569.25 + 4207.47] = 723.562$ which is highly significant at the 1% level. This provides evidence for the existence of incomplete information on both sides of the housing market. Table 2 presents the mean and quartile values of several measures based on the point predictors of $E(e^{-u}|\varepsilon)$ and $E(e^{-w}|\varepsilon)$, which are measures of sellers' and buyers' price efficiency. From these point predictors we find that, on average, buyers are 72% efficient and sellers are 70% efficient. That is, on average, buyers paid 28% above the lowest WTA, $(1-\hat{E}(e^{-w}|\varepsilon))$. The median and quartile values show large variations in information cost paid by the buyers. The last row of Table 3.2 shows that, on average, sellers received 30% less than the highest WTP, $(1-\hat{E}(e^{-u}|\varepsilon))$. The quartile values of sellers' incomplete information costs are quite similar to those of the buyers.

Table 3.1: Parameter Estimates of the Benchmark Hedonic Price Function*

Variable	Estimate	Variable	Estimate
Constant	6.8087 (*0.000*)	Rural	−0.3282 (*0.000*)
ln(Square Footage)	0.3731 (*0.000*)	Deemed Inadequate	−0.0555 (*0.392*)
Square Footage Top Coded	−0.1310 (*0.002*)	Degree Days Code (1-6)	−0.0128 (*0.014*)
Number of Bathrooms	0.3024 (*0.000*)	Sale Year:1987	0.0383 (*0.148*)
Number of Rooms	0.0302 (*0.000*)	Sale Year:1988	0.0835 (*0.002*)
Single Family Attached	0.7917 (*0.000*)	Sale Year:1989	0.1443 (*0.000*)
Single Family Detached	0.8599 (*0.000*)	Sale Year:1990	0.2136 (*0.000*)
Structure Age $\leq$ 5 years	0.2545 (*0.000*)	Sale Year:1991	0.1490 (*0.000*)
Structure Age 5-10 years	0.1402 (*0.000*)	Sale Year:1992	0.2073 (*0.000*)
Structure Age 10-15 years	0.0741 (*0.001*)	Sale Year:1993	0.2169 (*0.000*)
Structure Age $\geq$ 30 years	−0.0408 (*0.037*)	MSA $>$ 7 million	0.7063 (*0.000*)
Central City	−0.1074 (*0.000*)	MSA 3-7 million	0.2216 (*0.000*)
Urban/Suburban	0.0025 (*0.921*)	MSA 1-3 million	0.2289 (*0.000*)
Other Urban	−0.3047 (*0.000*)	not in MSA	−0.0844 (*0.001*)
σ_v	0.1454 (*0.000*)	σ_u	0.4288 (*0.000*)
		σ_w	0.3889 (*0.000*)

*The natural logarithm of the selling price is used as the dependent variable in the regression. Asymptotic p values are reported in parentheses next to each estimate. There are 4,962 observations.

Aside from investigating the price efficiency of individual buyers and sellers, one can also examine whether buyers are benefiting from sellers' incomplete information or vice-versa. Using the estimated values of $E(e^{-u}|\varepsilon)$ and $E(e^{-w}|\varepsilon)$ for each transaction, we construct an estimate of $100[e^{w-u}-1] = 100[e^{-u}/e^{-w}-1]$, which is the net effect of incomplete information on the price in percentage terms. Table 3.2 presents the mean and quartile values of the net information cost on price. The first row of Table 3.2 shows that in more than 50% of the transactions buyers benefited in the sense that the information costs paid by the sellers (which lowered price) exceeded those of the buyers (which increased price). The median net effect is −2.3%, thereby meaning that information deficiency led to a decline in house prices of 2.3% or more for over half the houses in our sample. For a quarter of the transactions the net effect was a reduction of price by at least 23.7%, while in another quarter of all transactions (those in the third quartile) the net information cost led to an increase in prices by 25.8% or more. These results show that the impact of incomplete information on housing prices is quite large and varied substantially across transactions (the inter-quartile range, i.e., the difference between the third and the first quartile values, is 49.5%).

Table 3.2: Cost of Incomplete Information on Housing Prices (Benchmark Model)*

Measure	Mean	Q_1	Median	Q_3	$Q_3 - Q_1$
$\hat{E}\left(\left(e^{w-u}\right) \mid \varepsilon\right) - 1$	11.6%	−23.7%	−2.3%	25.8%	49.4%
$1 - \hat{E}\left(e^{-w} \mid \varepsilon\right)$	28.1%	17.0%	19.9%	34.1%	17.1%
$1 - \hat{E}\left(e^{-u} \mid \varepsilon\right)$	30.0%	17.1%	21.7%	36.6%	19.5%

*Given that our dependent variable is in natural logarithms we use the point predictors of e^{-w} and e^{-u} to measure buyer's and seller's price efficiency.

Although, at the median, the information taxes imposed on buyers and sellers nearly cancel out, there remain some buyers and sellers who are benefitting from the other party's information deficiency. While we have considered differences between buyers and sellers, it also is of interest to investigate deficiency for different groups of buyers and sellers.

To explore the issue of differences in information deficiency across groups of buyers and sellers, we evaluate the costs of information deficiency across several groups. The results are presented in Tables 3.3 and 3.4 using point predictors of $1 - e^{-w}$ and $1 - e^{-u}$, which (when multiplied by 100) are the percentage increase/decrease in prices due to buyers'/sellers' information deficiency.

The entries in Table 3.3 represent the percent by which different types of buyers overpay relative to the minimum WTA. We find that, on average, the cost of information deficiency for first time home buyers and buyers with kids, are the least, while the cost to out of town buyers, blacks, single females, and those who own a business are somewhat higher. The first quartile values suggest almost no difference across different groups of buyers, while the third quartile values suggest that buyers who are from out of town, black or single female pay a higher tax for information deficiency. This result may be justified for those who own a business by arguing that their opportunity cost of time is higher. It is worth noting that substantial variation in the costs of incomplete information (measured by the interquartile range) are observed across different groups of buyers.

The entries in Table 3.4 represent the percent by which sellers' prices are reduced in relation to the maximum WTP (labeled as sellers' cost of incomplete information). On average, blacks and single females face the biggest price reductions (pay the highest information tax), while those who are college educated or own a business face a smaller tax. Comparing Tables 3.3 and 3.4, we find that the variation in incomplete information costs is much smaller for sellers in every category.

While these results are interesting, a general model that accounts for buyer's and seller's attributes directly into the hedonic price function, and indirectly through the means of u and w (i.e., σ_u and σ_w) is desirable. In this framework incomplete information costs are allowed to vary systematically across buyers and sellers. Results from this generalized model are reported next.

Table 3.3: Cost of Buyer's Information Deficiency (Benchmark Model)*

Buyer Type	Mean	Q_1	Median	Q_3	$Q_3 - Q_1$
Out of Town Buyer	30.0%	17.2%	22.5%	39.2%	22.0%
First Time Home Buyer	26.9%	17.0%	18.6%	30.7%	13.7%
Owns a Business	29.9%	17.7%	23.0%	37.7%	20.0%
Buyer is Black	30.1%	17.4%	21.4%	38.5%	21.1%
Buyer is Married	28.8%	17.1%	20.4%	35.7%	18.6%
Buyer is Single Female	29.9%	17.7%	23.2%	37.8%	20.1%
Buyer is College Educated	27.5%	17.0%	19.7%	33.0%	16.0%
Buyer has Kids	26.8%	17.0%	19.3%	31.7%	14.7%

*We use $1 - \hat{E}\left(e^{-w} \mid \varepsilon\right)$ to calculate all entries in the table.

Table 3.4: Cost of Seller's Information Deficiency (Benchmark Model)*

Seller Type	Mean	Q_1	Median	Q_3	$Q_3 - Q_1$
Owns a Business	27.1%	17.0%	20.1%	32.3%	15.2%
Seller is Black	39.4%	19.7%	33.0%	53.3%	33.6%
Seller is Married	28.4%	17.1%	21.4%	34.3%	17.2%
Seller is Single Female	33.0%	17.1%	23.8%	44.6%	27.5%
Seller is College Educated	25.4%	17.0%	19.1%	28.5%	11.6%
Seller has Kids	29.6%	17.3%	22.6%	36.7%	19.3%

*$1 - \hat{E}\left(e^{-u} \mid \varepsilon\right)$ is used to calculate all entries in the table.

3.5.3 Results from the Generalized Model

Table 3.5 reports the coefficient estimates of our linear hedonic price function, in which buyers' and sellers' attributes are incorporated both directly into the hedonic price function as well as indirectly through the means of u and w. The reason for doing so is to determine if direct inclusion of the supposed determinants of information render the use of the 2TSF useless. If we include these dummy variables and the variances of w and u dissipate, then we have a plausible argument for using the dummy variable approach, which is simpler, as opposed to the more complex frontier method. However, if we include these characteristic values and the variances still remain greater than zero, then we can conclude that the frontier method is capturing price fluctuations that are caused by incomplete information that a simple intercept shift cannot.

It can be seen from the Table 3.5 that the coefficients of the variables that were present in the benchmark model are nearly identical and their statistical significance is relatively unchanged. However, several of the buyer/seller attributes are important for explaining housing price variation.[14] For example, out of town buyers pay roughly 6.51% more than

[14]A likelihood ratio test failed to reject the null hypothesis that the buyer and seller attribute variables were jointly insignificant within the hedonic function specification.

Table 3.5: Parameter Estimates of the Extended Hedonic Price Function*

Variable	Estimate	Variable	Estimate
Constant	7.7839 (*0.000*)	Rural	−0.2801 (*0.000*)
Square Footage	0.2345 (*0.000*)	Deemed Inadequate	−0.0203 (*0.734*)
Square Footage Top Coded	−0.0736 (*0.027*)	Degree Days Code (1-6)	−0.0118 (*0.011*)
Number of Bathrooms	0.1786 (*0.000*)	Sale Year:1987	0.0266 (*0.246*)
Number of Rooms	0.0050 (*0.336*)	Sale Year:1988	0.0605 (*0.008*)
Single Family Attached	0.5637 (*0.000*)	Sale Year:1989	0.1142 (*0.000*)
Single Family Detached	0.6424 (*0.000*)	Sale Year:1990	0.1512 (*0.000*)
Structure Age $\leq$ 5 years	0.2041 (*0.000*)	Sale Year:1991	0.1021 (*0.000*)
Structure Age 5-10 years	0.0931 (*0.000*)	Sale Year:1992	0.1132 (*0.000*)
Structure Age 10-15 years	0.0258 (*0.184*)	Sale Year:1993	0.1400 (*0.000*)
Structure Age $\geq$ 30 years	−0.0258 (*0.123*)	MSA $>$ 7 million	0.5362 (*0.000*)
Central City	−0.1051 (*0.000*)	MSA 3-7 million	0.1715 (*0.000*)
Urban/Suburban	−0.0240 (*0.275*)	MSA 1-3 million	0.1927 (*0.000*)
Other Urban	−0.2929 (*0.000*)	not in MSA	−0.0567 (*0.010*)
Out of Town Buyer	0.0651 (*0.018*)	1st time Buyer	−0.0274 (*0.200*)
Buyer's Income	0.2314 (*0.000*)	Seller's Income	0.0571 (*0.000*)
Buyer has Business	0.0779 (*0.020*)	Seller has Business	0.1573 (*0.000*)
Buyer's Age	0.2020 (*0.000*)	Seller's Age	0.1736 (*0.000*)
Buyer is Black	−0.0544 (*0.167*)	Seller is Black	−0.0442 (*0.410*)
Buyer is Married	0.0515 (*0.056*)	Seller is Married	−0.0645 (*0.026*)
Buyer is Single Female	−0.0012 (*0.972*)	Seller is Single Female	0.0030 (*0.932*)
Buyer has College Education	0.1637 (*0.000*)	Seller has College Education	0.0414 (*0.020*)
Buyer has Kids	0.0220 (*0.250*)	Seller has Kids	−0.0220 (*0.264*)

*The natural logarithm of the selling price is used as the dependent variable in the regression. Asymptotic p values are reported in parentheses next to each estimate. The number of observations is 4,962.

a buyer from the local area, ceteris paribus. The coefficient for first time buyers is insignificant (at the 5% level of significance), suggesting that those who have bought a house before do not get a discount (Turnbull and Sirmans 1993). Other variables that are statistically insignificant are: blacks (both buyers and sellers), single females (both buyers and sellers), and families having kids (both buyers and sellers). Age, income, owning a business, and having a college education, are all found to be statistically significant and have the expected signs.

In Table 3.6 we report the estimate of σ_v and the parameters associated with the σ_u and σ_w functions. So far as buyers' characteristics are concerned, we find that only the black dummy is significant. The negative sign on the black dummy in σ_w suggests that the incomplete information costs for black buyers is lower (everything else being the same). However, given that only 2.4% of all buyers are black in our sample, this result may not be representative of the underlying population of black home buyers.

Table 3.6: Estimates of Parameters in the σ_u and σ_w Functions*

Variable	Estimate	Variable	Estimate
σ_v	0.1401 (*0.000*)		
σ_w		σ_u	
Constant	−1.1277 (*0.000*)	Constant	−0.4606 (*0.000*)
Out of Town Buyer	−0.0309 (*0.713*)		
1st Time Buyer	0.0287 (*0.655*)		
Buyer's Income	−0.0257 (*0.561*)	Seller's Income	−0.1721 (*0.000*)
Buyer has Business	−0.0362 (*0.637*)	Seller has Business	−0.0013 (*0.984*)
Buyer's Age	0.0965 (*0.330*)	Seller's Age	0.0064 (*0.933*)
Buyer is Black	−0.3537 (*0.0139*)	Seller is Black	0.2272 (*0.074*)
Buyer is Married	−0.1262 (*0.102*)	Seller is Married	−0.2787 (*0.000*)
Buyer is Single Female	0.0228 (*0.813*)	Seller is Single Female	−0.0030 (*0.700*)
Buyer has College Education	−0.0006 (*0.991*)	Seller has College Education	−0.2512 (*0.000*)
Buyer has Kids	−0.0896 (*0.123*)	Seller has Kids	−0.0655 (*0.233*)

*Asymptotic p values are reported in parentheses next to each estimate.

Next, we look at the sellers' side of the market. Unlike buyers, several variables are found to be significant in explaining σ_u. A seller's income, being married, and having a college education are all significant at the 1% level, while the black dummy is significant at the 10% level. All four of these characteristics have negative coefficients suggesting that the presence of these characteristics (given everything else) reduce the information cost paid by these sellers, on average. A likelihood ratio test failed to reject the null hypothesis that the buyer and seller attribute variables were jointly insignificant within the information tax distributions.

Again, as with the benchmark model, both variances incorporating incomplete information are individually significant. Again, a joint test of significance yields a likelihood ratio of $-2 \cdot [-4121.70 + 3508.60] = 1226.21$ which is highly significant at the 1% level and we have evidence that incomplete information exists in the housing market. Thus, the inclusion of buyer and seller attributes is not detracting from the main issue, that there are significant price fluctuations in the housing market due to less than full information on behalf of both buyers and sellers simultaneously.

As mentioned earlier, since both buyers and sellers are taxed due to incomplete information and one party's loss is other party's gain, it is more meaningful to examine the net effect of buyers and sellers information costs on prices. The results are reported in Table 3.7. Computations of all the measures reported here are the same as those in Table 3.2.

Table 3.7: Cost of Incomplete Information on Housing Prices (Extended Model)*

Measure	Mean	Q_1	Median	Q_3	$Q_3 - Q_1$
$\hat{E}\left((e^{w-u}) \mid \varepsilon\right) - 1$	1.3%	−23.9%	−4.6%	15.5%	39.4%
$1 - \hat{E}\left(e^{-w} \mid \varepsilon\right)$	23.3%	15.0%	17.5%	26.3%	11.3%
$1 - \hat{E}\left(e^{-u} \mid \varepsilon\right)$	28.1%	15.5%	20.4%	35.1%	19.6%

*Given that our dependent variable is in natural logarithms we use the point predictors of e^{-w} and e^{-u} to measure buyer's and sellers' price efficiency.

Similar to the benchmark model, we find that in more than 50% of the transactions buyers benefited in the sense that the tax on sellers exceeded that of the buyers, resulting in a price reduction. The median net effect shows a decline in price by 4.6% due to buyer and seller incomplete information. For a quarter of all transactions, the net effect was a reduction of price by at least 23.9%, while in another quarter of the transactions, net incomplete information costs led to an increase in prices by 15.5% or more. These results show that even after taking buyers' and sellers' characteristics into account within the hedonic price function (as has been done in previous research), the impact of buyers' and sellers' incomplete information on housing prices does not vanish. It is clear from Table 5 that variation in net incomplete information costs across transactions is quite large (the inter-quartile range is 39.5%), although somewhat smaller compared to the benchmark model (see Table 3.2).

In sum, the cost of information deficiency led to a price decrease (at the median) by 4.6%, as opposed to 2.3% in the benchmark model (reported in Table 3.2). This suggests that buyers, at the median, are benefiting from sellers' incomplete information. Intuitively this makes sense because buyers can search over a broad array of houses for very low cost, while sellers have to keep their house on the market longer if they wish to gather information from prospective buyers. Keeping a house on the market longer can result in a lower selling price as it may be evidence that the house is of poor quality (see Anglin (2005) for more on this).

Next we consider the cost of incomplete information across different types of buyers and sellers. The results are reported in Tables 3.8 (for buyers) and 3.9 (for sellers).

Table 3.8: Cost of Buyer's Information Deficiency (Extended Model)*

Buyer Type	Mean	Q_1	Median	Q_3	$Q_3 - Q_1$
Out of Town Buyer	22.9%	14.4%	17.4%	26.3%	11.8%
First Time Home Buyer	23.3%	15.1%	17.5%	26.3%	11.1%
Owns a Business	23.5%	14.9%	17.2%	27.2%	12.3%
Buyer is Black	18.0%	12.8%	14.5%	17.9%	5.1%
Buyer is Married	22.4%	14.5%	16.6%	25.4%	10.9%
Buyer is Single Female	25.8%	16.7%	18.9%	29.2%	12.5%
Buyer is College Educated	23.2%	14.3%	17.2%	27.8%	13.5%
Buyer has Kids	21.9%	14.4%	16.4%	24.8%	10.5%

*$1 - \hat{E}\left(e^{-w} \mid \varepsilon\right)$ is used to calculate all entries in the table.

It can be seen from Table 3.8 that differences in information taxes across groups of buyers disappear (except for the blacks) when the group characteristics are taken into account in the hedonic price function as well as in σ_u and σ_w. Like the benchmark model, the costs of information deficiency for black home buyers, *ceteris paribus*, is found to be lower.[15]

[15]Given that there are only 225 black buyers in our sample one must interpret this result with care.

Table 3.9: Cost of Seller's Information Deficiency (Extended Model)*

Seller Type	Mean	Q_1	Median	Q_3	$Q_3 - Q_1$
Owns a Business	26.6%	15.1%	20.2%	33.6%	18.5%
Seller is Black	33.2%	16.4%	23.6%	42.0%	25.7%
Seller is Married	25.2%	14.6%	18.6%	29.6%	15.1%
Seller is Single Female	35.7%	18.0%	28.6%	48.5%	30.5%
Seller is College Educated	22.7%	13.6%	16.9%	27.2%	13.7%
Seller has Kids	25.6%	14.5%	18.6%	30.4%	15.9%

*$1 - \hat{E}\left(e^{-u} \mid \varepsilon\right)$ is used to calculate all entries in the table.

Turning to the sellers in Table 3.9, we find that cost of information deficiency for blacks and single females are the highest (evidenced by the mean and quartile values). Costs of information deficiency for sellers with college education, having kids, being married, and owning a business are the least. Again this is evidenced by the mean and quartile values. Here the intuition is that sellers with the above characteristics can afford to leave their house on the market longer (thereby having more information about the prospective buyers) which brings down the cost of information deficiency. In summary, we find that the incomplete information for both buyers and sellers are nontrivial and these costs are quite similar for each group. Moreover, the variations in these costs for each group, measured by the inter-quartile range, are also similar.

3.6 Conclusion

In this paper we use a two-tier stochastic hedonic price function to decompose price variations into those explained by observed variables, unobserved product heterogeneity, and information deficiency. Since obtaining information is costly, and these costs are likely to vary among buyers and sellers, the presence of incomplete information imposes a tax/cost on both buyers and sellers. Such taxes may be optimal (efficient). That is, a buyer might pay a higher price for a good, knowing that the search costs (opportunity cost of time) associated with gathering additional information necessary for obtaining a lower price is too high. The same is true for a seller who might sell the good at a lower price instead of waiting longer to gather more information about prospective buyers who might be willing to pay more. The two-tier stochastic frontier approach used in this paper enabled us to obtain estimates of price efficiency and cost of information deficiency for each buyer and seller. We also extended the model to allow information costs to depend on buyers' and sellers' characteristics. This formulation allowed systematic variation in incomplete information costs, estimates of which are used to analyze differences in the cost of incomplete information across different types of buyers and sellers.

An application of the model to American Housing Survey data showed that the impact of incomplete information on market prices is not negligible for either buyers or sellers. On average, buyers were found to be 72% price the price efficiency of sellers was 70% efficient. That is, because of information deficiency buyers, on average, paid 28% above the lowest available WTA, while sellers, on average, received 30% less that the highest WTP. In the extended model that takes buyers' and sellers' characteristics into account, these figures are 23.3% and 28.1%, respectively. That is, information costs led to a net price decrease (at the median) of 4.6%, as opposed to 2.3% in the benchmark model. These results suggest that buyers, at the median, are benefiting from sellers' incomplete information.

The estimation of the costs of incomplete information may prove useful in many other economic settings where buyers (sellers) seek the lowest (highest) price and they may not share the same information. Some examples include price determination in auctions and used car markets, dowries in marriage markets, wages in labor markets, and premiums in insurance markets.

Bibliography

[1] Anderson Jr., R. J. and T. D. Crocker, 1972. "Air Pollution and Property Values: A Reply," *The Review of Economics and Statistics*, 54(4), pp. 470-473.

[2] Anglin, P. M., 2005. "The selling process: If, at first, you don't succeed, try, try, try again," Working Paper, University of Windsor.

[3] Anglin P. M. and R. Gençay, 1996. "Semiparametric Estimation of a Hedonic Price Function," *Journal of Applied Econometrics*, 11(6), pp. 633-48.

[4] Aitchison, J. and C. G. G. Aitken, 1976. "Multivariate Binary Discrimination by the Kernel Method," *Biometrika*, 63(3), pp. 413-420.

[5] Atkinson, S. E. and T.D. Crocker, 1987. "A Bayesian Approach to Assessing the Robustness of Hedonic Property," *Journal of Applied Econometrics*, 2(1), pp. 27-45.

[6] Bajari, P. and M. E. Kahn, 2005. "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities," *Journal of Business & Economic Statistics*, 23(1), pp. 20-33.

[7] Ball, M. J., 1973. "Recent Empirical Work on the Determinants of Relative House Prices," *Urban Studies*, 10(2), pp. 213-233.

[8] Baltagi, B. H., Y.-J. Chang, and Q. Li, 1992. "Monte Carlo Results on Several New and Existing Tests for the Error Component Model," *Journal of Econometrics*, 54(1), pp. 95-120.

[9] Bao, H. X. H. and A. T. K. Wan, 2004. "On the Use of Spline Smoothing in Estimating Hedonic Housing Price Models: Empirical Evidence Using Hong Kong Data," *Real Estate Economics*, 32(3), pp. 487-507.

[10] Bartik, T. J., 1987a. "The Estimation of Demand Parameters in Hedonic Price Models," *Journal of Political Economy*, 95(1), pp. 81-88.

[11] Bartik, T. J., 1987b. "Estimating Hedonic Demand Parameters with Single Market Data: The Problems Caused by Unobserved Tastes," *The Review of Economics and Statistics*, 69(1), pp. 178-80.

[12] Bartik, T. J., 1988. "Measuring the Benefits of Amenity Improvements in Hedonic Price Models," *Land Economics*, 64(2), pp. 72-83.

[13] Becker, G. S., 1965. "A Theory of the Allocation of Time," *Economic Journal*, 75(299), pp. 493-517.

[14] Bender, B., T. J. Gronberg, and H.-S. Hwang, 1980. "Choice of Functional Form and the Demand for Air Quality," *The Review of Economics and Statistics*, 62(4), pp. 638-643.

[15] Bin, O. 2004. "A Prediction Comparison of Housing Sales Prices by Parametric Versus Semi-Parametric Regression," *Journal of Housing Economics*, 13(1), pp. 68-84.

[16] Bin, O. 2005. "A Semiparametric Hedonic Model for Valuing Wetlands," *Applied Economics Letters*, 12(10), pp. 597-601.

[17] Blackley, P., J. R. Follain Jr., and J. Ondrich, 1984. " Box-Cox Estimation of Heodnic Models: How Serious is the Iterative OLS Variance Bias?" *The Review of Economics and Statistics*, 66(2), pp. 348-353.

[18] Blomquist, G. and L. Worley, 1981. "Hedonic Prices, Demands for Urban Housing Amenities, and Benefit Estimates," *Journal of Urban Economics*, 9(2), pp. 212-21.

[19] Blomquist, G. and L. Worley, 1982. "Specifying the Demand for Housing Characteristics: The Exogeneity Issue," in D. B. Diamond and G. S. Tolley (eds), *The Economics of Urban Amenities*, pp. 89-102, Academic Press: New York.

[20] Brown, J. N., 1983. "Structural Estimation in Implicit Markets," in J. Triplett (ed) *The Measurement of Labor Cost*, University of Chicago Press: Chicago.

[21] Brown, J. N. and H. S. Rosen, 1982. "On the Estimation of Structural Hedonic Price Models," *Econometrica*, 50(3), pp. 765-68.

[22] Can, A., 1992. "Specification and Estimation of Hedonic Housing Price Models," *Regional Science and Urban Economics*, 22(3), pp. 453-74.

[23] Cassel, E. and Mendelsohn, R., 1985. "The Choice of Functional Forms for Hedonic Price Equations: Comment," *Journal of Urban Economics*, 18(2), pp. 135-42.

[24] Chant, D., 1974. "On Asymptotic Tests of Composite Hypothesis in Nonstandard Conditions," *Biometrika*, 61(2), pp. 291-298.

[25] Chernoff, H., 1954. "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 25(3), pp. 573-578.

[26] Cheshire, P. and S. Sheppard, 1998. "Estimating the Demand for Housing, Land, and Neighbourhood Characteristics," *Oxford Bulletin of Economics and Statstics*, 60(3), pp. 357-382.

[27] Clapp, J. M., 2004. "A Semiparametric Method for Estimating Local House Price Indices," *Real Estate Economics*, 32(1), pp. 127-160.

[28] Clapp, J. M., H-J. Kim, and A. E. Gelfand, 2002. "Predicitng Spatial Patterns of House Prices Using LPR and Bayesian Smoothing," *Real Estate Economics*, 30(4), pp. 505-532.

[29] Coelli, T., 1995. "Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis," *The Journal of Productivity Analysis*, 6(3), pp. 247-268.

[30] Court, A. T., 1939. "Hedonic Price Indices with Automotive Examples," in *The Dynamics of Automobile Demand*, General Motors: New York.

[31] Court, L. M., 1941a. "Entrepreneurial and Consumer Demnd Theories for Commodity Spectra," *Econometrica*, 9(2), pp. 135-162.

[32] Court, L. M., 1941b. "Entrepreneurial and Consumer Demnd Theories for Commodity Spectra," *Econometrica*, 9(3), pp. 241-297.

[33] Cropper, M. L., L. B. Deck, and K. E. McConnell, 1988. "On the Choice of Functional Form for Hedonic Price Functions," *The Review of Economics and Statistics*, 70(4), pp. 668-75.

[34] de Leeuw, F., 1971. "The Demand for Housing: A Review of Cross-Section Evidence," *The Review of Economics and Statistics*, 53(1), pp. 1-10.

[35] Delgado, M. A. and W. González Manteiga, 2001. "Significance Testing in Nonparametric Regression Based on the Bootstrap," *Annals of Statistics*, 29(5), pp. 1469-1507.

[36] Diamond, D. B. Jr. and B. A. Smith, 1985. "Simultaneity in the Market for Housing Characteristics," *Journal of Urban Economics*, 17(3), pp. 280-292.

[37] Ekeland, I., J. J. Heckman, and L. Nesheim, 2002. "Identifying Hedonic Models," *The American Economic Review: Papers and Proceedings*, 92(2), pp. 304-309.

[38] Ekeland, I., J. J. Heckman, and L. Nesheim, 2004. "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1), pp. S60-S109.

[39] Epple, D., 1987. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *Journal of Political Economy*, 95(1), pp. 59-80.

[40] Fan, J., 1992. "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87(420), pp. 998-1004.

[41] Fan, J., 1993. "Local Linear Regression Smoothers and Their Minimax Efficiencies," *Annals of Statistics*, 21(1), pp. 196-216.

[42] Follain, J. and E. Jimenez, 1985. "Estimating the Demand for Housing Characteristics: A Survey and Critique," *Regional Science and Urban Economics*, 15(1), pp. 77-107.

[43] Freeman, A. M. III, 1971. "Air Pollution and Property Values: A Methodological Comment," *The Review of Economics and Statistics*, 53(4), pp. 415-416.

[44] Freeman, A. M. III, 1974a. "Air Pollution and Property Values: A Further Comment," *The Review of Economics and Statistics*, 56(4), pp. 554-556.

[45] Freeman, A. M. III, 1974b. "On Estimating Air Pollution Control Benefits from Land Value Studies," *Journal of Environmental Economics and Management*, 1(2), pp. 74-83.

[46] Freeman, A. M. III, 1979. "Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues," *Scandinavian Journal of Economics*, 81(2), pp. 154-73.

[47] Gençay, R. and X. Yang, 1996. "A Forecast Comparison of Residential Housing Prices by Parametric versus Semiparametric Conditional Mean Estimators," *Economics Letters*, 52(2), pp. 129-35.

[48] Goodman, A. C. 1978. "Hedonic Prices, Price Indices and Housing Markets," *Journal of Urban Economics*, 5(4), pp. 471-84.

[49] Gordon, R. J. 1973. "The Measurement of Durable Goods Prices," NBER, mimeo.

[50] Gourieroux, C., A. Holly, and A. Monfort, 1982. "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, 50(1), pp. 63-80.

[51] Graves, P., J. C. Murdoch, M. A. Thayer, and D. Waldman, 1988. "The Robustness of Hedonic Price Estimation: Urban Air Quality," *Land Economics*, 64(3), pp. 220-33.

[52] Griliches, Z. 1961. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," *Government Price Statistics, Hearings*, U. S. Congress, Joint Economic Commitee, pp. 173-196.

[53] Griliches, Z. 1971. *Price Indices and quality Change*. Harvard University Press: Cambridge, MA.

[54] Halvorsen, R. and H. O. Pollakowski, 1981. "Choice of Functional Form for Hedonic Price Equations," *Journal of Urban Economics*, 10(1), pp. 37-49.

[55] Harding, J. P., S. S. Rosenthal, and C. F. Sirmans, 2003. "Estimating Bargaining Power in the Market for Existing Homes," *The Review of Economics and Statistics*, 85(1), pp. 178-188.

[56] Härdle, W. K., 1990. *Applied Nonparametric Regression*, Cambridge University Press: Boston.

[57] Harrison, D. and Rubinfeld D.L. 1978a. "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, 5(1), pp. 81-102.

[58] Harrison, D. and Rubinfeld D. L. 1978b. "The Air Pollution and Property Value Debate: Some Empirical Evidence," *The Review of Economics and Statistics*, 60(4), pp. 635-638.

[59] Hartog, J. and H. Bierens, 1991. "Estimating a Hedonic Earnings Function with a Nonparametric Method," in A. Ullah (ed.), *Semiparametric and Nonparametric Econometrics: Studies in Empirical Economics*, Springer: Berlin, Heidelberg, New York.

[60] Heckman, J. J., R. Matzkin, and L. Nesheim, 2002. "Identification and Estimation of Hedonic Models: The Vector Nonseparable Case with Missing Attribute," Manuscript. Chicago: University of Chicago, Department of Economics.

[61] Heckman, J. J., R. Matzkin, and L. Nesheim, 2003. "Simulation and Estimation of Nonadditive Hedonic Models," NBER Working Paper No. 9895.

[62] Henderson, D. J. and D. L. Millimet, 2006. "Is Gravity Linear?" Working Paper, Southern Methodist Univeristy, Department of Economics.

[63] Honda, Y., 1985. "Testing the Error Components Model with Non-Normal Disturbances," *Review of Economic Studies*, 52(4), pp. 681-690.

[64] Horowitz, J. L., 1987. "Identification and Stochastic Specification in Rosen's Hedonic Price Model," *Journal of Urban Economics*, 22(2), pp. 165-73.

[65] Houthakker, H. S. 1952. "Compensated Changes in Quantities and Qualities Consumed," *Review of Economic Studies*, 19(3), 155-164.

[66] Hurvich, C. M., J. S. Simonoff, and C-L. Tsai, 1998. "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion," *Journal of the Royal Statistical Society Series B*, 60(2), pp. 271-93.

[67] Iwata, S., H. Murao, and Q. Wang, 2000. "Nonparametric Assessment of the Effects of Neighborhood Land Uses on the Residential House Values," in T. Fomby and R. Carter Hill (eds.), *Advances in Econometrics Volume 14: Applying Kernel and Nonparametric Estimation to Economic Topics*, JAI Press: New York.

[68] Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt, 1982, "On the estimation of technical inefficiency in the stochastic frontier production function model," *Journal of Econometrics*, 19(2/3), pp. 233-238.

[69] Kahn, S. and K. Lang, 1988. "Efficient Estimation of Structural Hedonic Systems," *International Economic Review*, 29(1), pp. 157-66.

[70] Kain, J. F. and J. M. Quigley, 1970. "Measuring the Value of Housing Quality," *Journal of the American Statistical Association*, 65(330), pp. 532-548.

[71] Kanemoto, Y. and R. Nakamura, 1986. "A New Approach to the Estimation of Structural Equations in Hedonic Models," *Journal of Urban Economics*, 19(2), pp. 218-33.

[72] Kask, S. B. and S. A. Maani, 1992. "Uncertainty, Information, and Hedonic Pricing," *Land Economics*, 68(2), pp. 170-84.

[73] Kim, S., 1992. "Search, Hedonic Prices and Housing Demand," *The Review of Economics and Statistics*, 74(3), pp. 503-08.

[74] Kumbhakar, S. C. and C. A. K. Lovell, 2000. *Stochastic Frontier Analysis*, Cambridge University Press: New York.

[75] Kumbhakar, S. C. and C. F. Parmeter, 2005. "The effects of bargaining on market outcomes: Evidence from buyer and seller specific estimates," Working Paper, State University of New York- Binghamton.

[76] Kumbhakar, S. C. and C. F. Parmeter, 2006. "Estimation of a Hedonic Price Function with Incomplete Information," Working Paper, State University of New York- Binghamton.

[77] Lancaster, K. J., 1966. "A New Approach to Consumer Theory," *Journal of Political Economy*, 74(2), 132-56.

[78] Lee, J., S-J. Kwak, and J. A. List, 2000. "Average Derivative Estimation of Hedonic Price Models," *Environmental and Resource Economics*, 16(1), pp. 81-91.

[79] LeSage, J. P. and R. K. Pace, 2004. *Advances in Econometrics, Volume 18: Spatial and Spatiotemporal Econometrics*, Elsevier Science: Amsterdam.

[80] Li Q. and J. S. Racine, 2004. "Cross-Validated Local Linear Nonparametric Regression," *Statistica Sinica*, 14(2), pp. 485-512.

[81] Linneman, P., 1980. "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market," *Journal of Urban Economics*, 8(1), pp. 47-68.

[82] Linneman, P., 1981. "The Demand for Residence Site Characteristics," *Journal of Urban Economics*, 9(1), pp. 129-148.

[83] Martins-Filho, C. and O. Bin, 2005. "Estimation of Hedonic Price Functions via Additive Nonparametric Regression," *Empirical Economics*, 30(1), pp. 93-114.

[84] Mason, C. and J. M. Quigley, 1996. "Non-parametric Hedonic Housing Prices," *Housing Studies*; 11(3), pp. 373-85.

[85] Mas-Colell, A., 1975. "A Model of Equilibrium with Differentiated Commodities," *Journal of Mathematical Economics*, 2(2), pp. 263-295.

[86] McConnell, K. E. and T. T. Phipps, 1987. "Identification of Preference Parameters in Hedonic Models: Consumer Demands with Nonlinear Budgets," *Journal of Urban Economics*, 22(1), pp. 35-52.

[87] McMillen, D. P. and P. Thorsnes, 2000. "The Reaction of Housing Prices to Information of Superfund Sites," in T. Fomby and R. Carter Hill (eds.), *Advances in Econometrics Volume 14: Applying Kernel and Nonparametric Estimation to Economic Topics*, JAI Press: New York.

[88] Meese, R. A. and N. E. Wallace, 1991. "Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices," *Journal of the American Real Estate and Urban Economics Association*, 19(3), pp. 308-333.

[89] Meese, R. A. and N. E. Wallace, 1997. "The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches," *Journal of Real Estimate Finance and Economics*, 14(1), pp. 51-73.

[90] Mendelsohn, R., 1984. "Estimating the Structural Equations of Implicit Markets and Household Production Functions," *The Review of Economics and Statistics*, 66(4), pp. 673-677.

[91] Mendelsohn, R., 1985. "Identifying Structural Equations with Single Market Data," *The Review of Economics and Statistics*, 67(3), pp. 525-529.

[92] Mendelsohn, R., 1987. "A Review of Identification of Hedonic Supply and Demand Functions," *Growth and Change*, 18(1), pp. 82-92.

[93] Miller, J. J., 1977. "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," *The Annals of Statistics*, 5(4), pp. 746-762.

[94] Moran, P. A. P., 1971. "Maximum Likelihood Estimators in Non-Stnadard Conditions," *Proceedings of the Cambridge Philosophical Society*, 70(2), pp. 441-450.

[95] Murray, M. P., 1983. "Mythical Demands and Mythical Supplies for Proper Estimation of Rosen's Hedonic Price Model," *Journal of Urban Economics*, 14(3), pp. 326-37.

[96] Muth, R. F., 1966. "Household Production and Consumer Demand Functions," *Econometrica*, 34(3), pp. 699-708.

[97] N©, Nonparametric software by Jeff Racine (http://www.economics.mcmaster.ca/racine/).

[98] Nelson, J. P. 1978. "Residential Choice, Hedonic Prices, and the Demand for Urban Air Quality," *Journal of Urban Economics*, 5(3), pp. 357-69.

[99] Ohsfeldt, R. L. and B. A. Smith, 1985. "Estimating the Demand for Hetergeneous Goods," *The Review of Economics and Statistics*, 67(1), pp. 165-171.

[100] Pagan A., and A. Ullah, 1999. *Nonparametric Econometrics*, Cambridge University Press: Cambridge, UK.

[101] Pace, R. K., 1993. "Nonparametric Methods with Applications to Hedonic Models," *Journal of Real Estate Finance and Economics*, 7(3), pp. 185-204.

[102] Pace, R. K., 1995. "Parametric, Semiparametric, and Nonparametric Estimation of Characteristic Values within Mass Assessment and Hedonic Pricing Models," *Journal of Real Estate Finance and Economics*, 11(3), pp. 195-217.

[103] Pace, R. K., 1998. "Appraisal Using Generalized Additive Models," *The Journal of Real Estate Research*, 15(1/2), pp. 77-99.

[104] Palmquist, R. B., 1984. "Estimating the Demand for the Characteristics of Housing," *The Review of Economics and Statistics*, 66(3), pp. 394-404.

[105] Palmquist, R. B., 1991. "Hedonic Methods," in J. B. Braden and C.D. Kolstad, (eds.), *Measuring the Demand for Environmental Quality*, Elsevier Scince: Amsterdam.

[106] Parmeter, C. F., D. J. Henderson, and S. C. Kumbhakar. "Nonparametric Estimation of a Hedonic Price Function," *Journal of Applied Econometrics*, forthcoming.

[107] Polachek, S. and B. J. Yoon, 1987. "A Two-Tiered Earnings Frontier Estimation of Employer and Employee Information in the Labor Market," *The Review of Economics and Statistics*, 69(2), pp. 296-302.

[108] Polachek, S. and B. J. Yoon, 1996. "Panel Estimates of a Two-Tiered Earnings Frontier," *Journal of Applied Econometrics*, 11(2), pp. 169-178.

[109] Polinsky, A. M. and S. Shavell, 1975. "The Air Pollution and Property Value Debate," *The Review of Economics and Statistics*, 57(1), pp. 100-104.

[110] Pratt, J., D. Wise, and R. Zeckhauser, 1979. "Price Differences in Almost Competitive Markets," *Quarterly Journal of Economics*, 93(2), pp. 189-211.

[111] Quigley, J. M., 1982. "Nonlinear Budget Constraints and Consumer Demand: An Application to Public Programs for Residential Housing," *Journal of Urban Economics*, 27(3), pp. 177-201.

[112] Racine J. S. and Q. Li, 2004. "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics*, 119(1), pp. 99-130.

[113] Rasmussen, D. W., and T. W. Zuehlke, 1990. "On the Choice of Functional Form for Hedonic Price Functions," *Applied Economics*, 22(4), pp. 431-38.

[114] Ridker, R. G. and J. A. Henning, 1967. "The Determinants of Residential Property Values with Special References to Air Pollution," *The Review of Economics and Statistics*, 49(2), pp. 246-257.

[115] Robinson, P. M., 1988. "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), pp. 931-54.

[116] Rosen, S., 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), pp. 34-55.

[117] Rothschild, M., 1973. "Models of Market Organization with Imperfect Information: A Survey," *Journal of Political Economy* 81(6), pp. 1283-1308.

[118] Self, S. G. and K.-Y. Liang, 1987. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *Journal of the American Statistical Society*, 82(398), pp. 605-610.

[119] Silverman B. W., 1986. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall: London.

[120] Small, K. A., 1975. "Air Pollution and Property Values: Further Comment," *The Review of Economics and Statistics*, 57(1), pp. 105-107.

[121] Stigler, G. 1961. "The Economics of Information," *Journal of Political Economy*, 69(3), pp. 213-225.

[122] Stock, J. 1989. "Nonparametric Policy Analysis," *Journal of the American Statistical Association* 84(406), pp. 567-75.

[123] Stock, J. 1991. "Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits", in W. Barnett, J. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of*

the Fifth International Symposium in Economic Theory and Econometrics, Cambrdige University Press: New York.

[124] Tinbergen, J. 1956. "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, 77(2), pp. 155-173.

[125] Turnbull, G. and C. F. Sirmans, 1993. "Information, Search and House Prices," *Regional Science and Urban Economics* 23(4), pp. 545-557.

[126] Wallace, N. E., 1996. "Hedonic-Based Price Indices for Housing: Theory, Estimation, and Index Construction," *Economic Review - Federal Reserve Bank of San Francisco*, 3(2), pp. 34-48.

[127] Wang, H. and P. Schmidt, 2002, "One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels," *Journal of Productivity Analysis* 18(2), pp. 129-144.

[128] Wang, M-C. and J. van Ryzin, 1981. "A Class of Smooth Estimators for Discrete Distributions," *Biometrika*, 68(1), pp. 301-309.

[129] Waugh, F. V., 1929. *Quality as a Determinant of Vegetable Prices*, Columbia University Press: New York.

[130] Witte, A. D., H. J. Sumka, and H. Erekson, 1979. "An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets," *Econometrica*, 47(5), pp. 1151-73.

Appendix A

Derivations Needed for Two-Tier Frontier Estimation

Derivation[1] of equation (3.8):

Beginning with the definition of the composed error term $\varepsilon_1 = v - u$, the marginal distribution of this is, following Kumbhakar and Lovell (2001),

$$f(\varepsilon_1) = (1/\sigma_u)\left(\Phi\left(-\varepsilon_1/\sigma_v - \sigma_v/\sigma_u\right)\exp\left\{\varepsilon_1/\sigma_u + \sigma_v^2/2\sigma_u^2\right\}\right). \tag{A.1}$$

The three component error may then be written as $\varepsilon = \varepsilon_1 + w$, which implies that $\varepsilon_1 = \varepsilon - w$, yielding the following joint distribution, $g(\varepsilon, w) = g(\varepsilon_1, w) \cdot |d\varepsilon_1/d\varepsilon| = g(\varepsilon_1, w) = f(\varepsilon_1) \cdot f(w)$. Upon integrating out w one obtains the marginal distribution of ε . This is done below.

$$\begin{aligned}
f(\varepsilon) &= \int_0^\infty \frac{1}{\sigma_u}\left(\Phi\left(-\frac{\varepsilon_1}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)\exp\left\{\frac{\varepsilon_1}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right\}\right)\frac{1}{\sigma_w}\exp\left\{-\frac{w}{\sigma_w}\right\}\, dw \\
&= \frac{1}{\sigma_u\sigma_w}\left[\exp\left\{\frac{\varepsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right\}\int_0^\infty \Phi\left(\frac{w}{\sigma_v} - \left(\frac{\varepsilon}{\sigma_v} + \frac{\sigma_v}{\sigma_u}\right)\right)\exp\left\{-w\left(\frac{1}{\sigma_w} + \frac{1}{\sigma_u}\right)\right\}\, dw\right] \\
&= \frac{-1}{\sigma_u + \sigma_w}\left[\exp\left\{\frac{\varepsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right\}\int_0^\infty \Phi\left(\frac{w}{\sigma_v} - \left(\frac{\varepsilon}{\sigma_v} + \frac{\sigma_v}{\sigma_u}\right)\right)\, d\left(\exp\left\{-w\left(\frac{1}{\sigma_w} + \frac{1}{\sigma_u}\right)\right\}\right)\right] \\
&= \frac{-\exp\{\alpha\}}{\sigma_u + \sigma_w}\left[\Phi(w/\sigma_v + \beta)\exp\{-w\lambda\}|_0^\infty - \int_0^\infty \phi(w/\sigma_v + \beta)\exp\{-w\lambda\}\, dw\right] \\
&= \frac{\exp\{\alpha\}}{\sigma_u + \sigma_w}\left[\Phi(\beta) + \exp\{-\alpha\}\exp\left\{\frac{\sigma_v^2}{2\sigma_w^2} - \frac{\varepsilon}{\sigma_w}\right\}\int_0^\infty \frac{1}{\sigma_v}\phi\left(\frac{w}{\sigma_v} - \left(\frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_w}\right)\right)\, dw\right] \\
&= \frac{\exp\{\alpha\}}{\sigma_u + \sigma_w}\Phi(\beta) + \frac{\exp\{a\}}{\sigma_u + \sigma_w}\int_{-b}^\infty \phi(z)\, dz = \frac{\exp\{\alpha\}}{\sigma_u + \sigma_w}\Phi(\beta) + \frac{\exp\{a\}}{\sigma_u + \sigma_w}\Phi(b)
\end{aligned} \tag{A.2}$$

[1] To avoid notational clutter the i subscript has been dropped in all of the derivations.

where $\lambda = \frac{1}{\sigma_u} + \frac{1}{\sigma_w}$, $\alpha = \frac{\varepsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}$, $\beta = -\left(\frac{\varepsilon}{\sigma_v} + \frac{\sigma_v}{\sigma_u}\right)$, $a = \frac{\sigma_v^2}{2\sigma_w^2} - \frac{\varepsilon}{\sigma_w}$, and $b = \frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_w}$.

To derive equations (3.12) and (3.13) we first need the conditional distributions of u and w, which is done as follows:

$$\begin{aligned} f(u|\varepsilon) &= \frac{f(u,\ \varepsilon)}{f(\varepsilon)} = \frac{(\exp\{a\}/\sigma_u\sigma_w)\exp\{-\lambda u\}\Phi(u/\sigma_v + b)}{(1/(\sigma_u+\sigma_w))[\exp\{a\}\Phi(b) + \exp\{\alpha\}\Phi(\beta)]} \\ &= \frac{\lambda\exp\{a\}\exp\{-\lambda u\}\Phi(u/\sigma_v + b)}{[\exp\{a\}\Phi(b) + \exp\{\alpha\}\Phi(\beta)]} \\ &= \frac{\lambda\exp\{-\lambda u\}\Phi(u/\sigma_v + b)}{\chi_1} \end{aligned} \tag{A.3}$$

where $\chi_1 = \Phi(b) + \exp\{\alpha - a\}\Phi(\beta)$.

Similarly,

$$\begin{aligned} f(w|\varepsilon) &= \frac{f(w,\ \varepsilon)}{f(\varepsilon)} = \frac{(\exp\{\alpha\}/\sigma_u\sigma_w)\exp\{-\lambda w\}\Phi(w/\sigma_v + \beta)}{(1/(\sigma_u+\sigma_w))[\exp\{a\}\Phi(b) + \exp\{\alpha\}\Phi(\beta)]} \\ &= \frac{\lambda\exp\{\alpha\}\exp\{-\lambda w\}\Phi(w/\sigma_v + \beta)}{[\exp\{a\}\Phi(b) + \exp\{\alpha\}\Phi(\beta)]} \\ &= \frac{\lambda\exp\{-\lambda w\}\Phi(w/\sigma_v + \beta)}{\chi_2} \end{aligned} \tag{A.4}$$

where $\chi_2 = \Phi(\beta) + \exp\{a - \alpha\}\Phi(b) = \exp\{a - \alpha\}\chi_1$.

Given these conditional distributions, our derivations of equations (3.12) and (3.13) are as follows:

$$\begin{aligned} E(e^{-u}|\varepsilon) &= \int_0^\infty e^{-u}\frac{\lambda e^{-\lambda u}\Phi(u/\sigma_v + b)}{\chi_1}\ du = \frac{\lambda}{\chi_1}\int_0^\infty e^{-(1+\lambda)u}\Phi(u/\sigma_v + b)\ du \\ &= \left(\frac{-\lambda}{\chi_1(1+\lambda)}\right)\int_0^\infty \Phi(u/\sigma_v + b)\ d(e^{-(1+\lambda)u}). \end{aligned} \tag{A.5}$$

Using integration by parts, we obtain

$$E\left(e^{-u}\left|\varepsilon\right.\right)=\left(\frac{-\lambda}{\chi_1\left(1+\lambda\right)}\right)\left[\Phi\left(u/\sigma_v+b\right)e^{-(1+\lambda)u}\Big|_0^\infty-\int_0^\infty e^{-(1+\lambda)u}\phi\left(u/\sigma_v+b\right)\,du/\sigma_v\right]$$
$$=\left(\frac{\lambda}{\chi_1\left(1+\lambda\right)}\right)\left[\Phi\left(b\right)+e^{\alpha-a+.5\sigma_v^2-\sigma_v\beta}\int_0^\infty\phi\left(u/\sigma_v+\left(b+\sigma_v\left(1+\lambda\right)\right)\right)\,du/\sigma_v\right].\tag{A.6}$$

Using the change of variable, $z=\frac{u}{\sigma_v}+\left(b+\sigma_v\left(1+\lambda\right)\right)\Rightarrow dz=du/\sigma_v$, we have

$$E\left(e^{-u}\left|\varepsilon\right.\right)=\left(\frac{\lambda}{\chi_1\left(1+\lambda\right)}\right)\left[\Phi\left(b\right)+e^{\alpha-a+.5\sigma_v^2-\sigma_v\beta}\Phi\left(\beta-\sigma_v\right)\right].\tag{A.7}$$

Similarly for $E(e^{-w}|\varepsilon)$,

$$E\left(e^{-w}\left|\varepsilon\right.\right)=\int_0^\infty e^{-w}\frac{\lambda e^{-\lambda w}\Phi\left(w/\sigma_v+\beta\right)}{\chi_2}\,dw=\left(\lambda/\chi_2\right)\int_0^\infty e^{-(1+\lambda)w}\Phi\left(w/\sigma_v+\beta\right)\,dw$$
$$=\left(\frac{-\lambda}{\chi_2\left(1+\lambda\right)}\right)\int_0^\infty\Phi\left(w/\sigma_v+\beta\right)\,de^{-(1+\lambda)w}.\tag{A.8}$$

Using integration by parts

$$E\left(e^{-w}\left|\varepsilon\right.\right)=\left(\frac{-\lambda}{\chi_2\left(1+\lambda\right)}\right)\left[\Phi\left(w/\sigma_v+\beta\right)e^{-(1+\lambda)w}\Big|_0^\infty-\int_0^\infty e^{-(1+\lambda)w}\phi\left(w/\sigma_v+\beta\right)\,dw/\sigma_v\right]$$
$$=\left(\frac{\lambda}{\chi_2\left(1+\lambda\right)}\right)\left[\Phi\left(\beta\right)+e^{a-\alpha-b\sigma_v+.5\sigma_v^2}\int_0^\infty\phi\left(\frac{w}{\sigma_v}+\left(\beta+\sigma_v\left(1+\lambda\right)\right)\right)\,dw/\sigma_v\right].\tag{A.9}$$

Finally, using the change of variable, $z=\frac{w}{\sigma_v}+\left(\beta+\sigma_v\left(1+\lambda\right)\right)\Rightarrow dz=dw/\sigma_v$, we have

$$E\left(e^{-w}\left|\varepsilon\right.\right)=\left(\frac{\lambda}{\chi_2\left(1+\lambda\right)}\right)\left[\Phi\left(\beta\right)+e^{a-\alpha-b\sigma_v+.5\sigma_v^2}\,\Phi\left(b-\sigma_v\right)\right].\tag{A.10}$$