IDENTIFICATION OF PREFERENCES IN HEDONIC MODELS

Volume I

of

BENEFIT ANALYSIS USING
INDIRECT OR IMPUTED MARKET METHODS

Prepared and Edited by

Kenneth E. McConnell
University of Maryland

Maureen Cropper
University of Maryland

Robert Mendelsohn
Yale University

Tim T. Phipps
Resources for the Future

Principal Investigators

Nancy E. Bockstael and Kenneth E. McConnell
Agricultural and Resource Economics
University of Maryland

EPA Contract No. CR-811043-01-0

Project Officer

Dr. Alan Carlin
Office of Policy Analysis
Office of Policy and Resource Management
U. S. Environmental Protection Agency
Washington, D. C. 20460

IDENTIFICATION OF PREFERENCES IN HEDONIC MODELS

Volume I

of

BENEFIT ANALYSIS USING INDIRECT OR IMPUTED MARKET METHODS

Prepared and Edited by

Kenneth E. McConnell
University of Maryland

Maureen Cropper
University of Maryland

Robert Mendelsohn
Yale University

Tim T. Phipps
Resources for the Future

Principal Investigators

Nancy E. Bockstael and Kenneth E. McConnell
Agricultural and Resource Economics
University of Maryland

EPA Contract No. CR-811043-01-0

Project Officer

Dr. Alan Carlin
Office of Policy Analysis
Office of Policy and Resource Management
U. S. Environmental Protection Agency
Washington, D. C. 20460

"The information in this document has been funded wholly or in part by the United States Environmental Protection Agency under Contract No. 811043-01-0. It has been subject to the Agency's peer and administrative review, and it has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use."

PREFACE

This report on identification in hedonic models represents the first year's work on the hedonic portion of the Cooperative Agreement between EPA and the University of Maryland. It will be followed by additional work on hedonics which investigates more fully the empirical issues associated with using the hedonic model to value environmental amenities.

In addition to the authors, a number of other people contributed to the ideas of this report. Both Kerry Smith and Michael Hanemann were influential in the development of Chapters 4 and 6.

Thorough review of reports is a characteristic of EPA Cooperative Agreements. This report benefited from the detailed comments and criticisms of the following individuals:

Raymond Palmquist
North Carolina State University

George Parsons
Environmental Protection Agency

Walter Milon
University of Florida
(on leave at EPA at the time of the review)

A number of graduate students helped draw figures, proofread, and otherwise assist in the preparation of the report. They include Douglas Orr, Terry Smith, Bruce Madariaga, Utpal Vasavada, Chester Hall and Laurence Crane.

Our contract officers on the research, Alan Carlin and Peter Caulkins, have been supportive and patient.

Finally, it is worth noting that this report represents the initial year's work on hedonics in a Cooperative Agreement that is designed to last four years. Additional work now under way will confront the conceptual questions with numbers.

EXECUTIVE SUMMARY

IDENTIFICATION OF PREFERENCES IN HEDONIC MODELS

EPA Cooperative Agreement CR-811043-01-0
University of Maryland, College Park, Maryland

Volume I

N. E. Bockstael and K. E. McConnell
Principal Investigators

This volume reports on the research of our project under the EPA Cooperative Agreement with the University of Maryland. The purpose of this project is "to solve the identification problem in hedonic models." The purpose of the research is thus quite specific and rather theoretical in nature. This volume describes those circumstances under which the problem is solved and analyzes other issues consistent with the use of the hedonic model in benefit-cost analysis.

The results of the project, while relating to technical issues, can be expressed intuitively. The hedonic model is a method of assessing the economic costs of pollution. Its use in environmental economics stems from the fact that when people buy homes, their willingness to pay for the attributes of the house is reflected in the sale price. The attributes of the house include not only its size and number of rooms, but also neighborhood characteristics and various dimensions of environmental quality, including air quality. Hedonic analysis connotes various approaches to the empirical study of the price of goods, when those prices reflect the characteristics of goods. For example, consider two houses which are located next to one another and differ only in that one house has an extra bathroom. Then when the housing market is in equilibrium, the difference in the housing prices reflects the additional bathroom. This basic principle allows us to impute housing price differences to differences in several attributes of houses, including environmental quality. Further, we can say the difference in the home price reflects a household's willingness to pay for the attribute. Consider two houses identical except that one is located in [illegible] The difference in the home prices reflects a household's willingness to pay for reductions in ozone.

The identification problem concerns the difficulties researchers encounter in trying to find the household's schedule of willingness to pay for various levels of attributes, not just a small change in the attribute. The identification problem stems from the fact that observed hedonic prices reflect not only on the value of the attribute to the household but also on the distribution of households of various types, the scarcity of houses, and the distribution of housing characteristics in the stock of housing.

In the context of benefit-cost analysis, the identification problem makes it more difficult to infer the benefits of non-marginal changes in attributes. Hedonic prices show what households would pay for small changes in housing traits, not their schedules of willingness to pay for various levels of the attributes. In measuring the value of various kinds of goods and service in the economy, we typically find that the more of a good a person has, the less he would be willing to pay for additional units of the good. Consequently, it would be wrong to compute how much a person would pay for 10 gallons of milk per week by finding what he pays for one gallon and multiplying by 10. The same holds for attributes of houses, including environmental attributes. The solution to the identification problem would therefore permit more accurate measurement of the benefits of the non-marginal changes in environmental amenities reflected in housing prices.

The basic finding concerning the solution to the identification problem when housing prices come from only one housing market is negative. Chapter 3 and 4 address the issue in detail. These chapters differ in how they address the problem, but both demonstrate that identification of the household's functional relationship between attribute levels and willingness to pay can be achieved only when the hedonic prices obey curvature patterns significantly different from the curvature of the individual willingness to pay function. Further, it is shown that the curvature properties which permit identification are not testable, but must simply be assumed. We are therefore in a position of solving the identification problem, but of not being able to test whether households behave in a way compatible with the assumptions that allow identification.

When we combine housing prices from different markets, for example, from different cities, the situation is not quite so pessimistic. If we are willing to believe without testing that households from different cities value attributes of houses approximately the same, then we may be able to identify the hedonic model (Chapter 4, Section 4).

Is it worthwhile to proceed with attempts to identify hedonic models? The answer depends on several factors. First, can we be satisfied that housing markets work approximately as hedonic analysis specifies? Second, does the estimation of the hedonic price equation--the relationship between housing prices and housing attributes--give an accurate reflection of what is going on in the housing market? Third, are there serious damages using [illegible]

Chapters 5 though 7 explore these issues. Chapter 5 asks whether the identification problem which plagues the recovery of information about willingness to pay for environmental attributes also confuses us about the term hedonic price equation. The answer is basically no.

Chapter 6 explores how much difference it makes to use marginal prices to calculate the benefits of non-marginal changes. The conclusion is that errors from using marginal prices are less serious than errors from other sources, such as specification of the hedonic relationship.

Chapter 7 investigates the structure of choice in hedonic models. It recognizes that residential locational choice can be viewed as a choice of two dimensions on a plane. If air pollution is tied systematically to either or both of these dimensions, then differences in housing prices will not reflect differences in willingness to pay for tied attributes. This chapter suggests that we may achieve more reliable results for the economic costs of pollution by developing a more realistic model of individual bids.

The conclusion of this volume is that while it is conceptually possible to identify the hedonic model, it is not a good use of research resources. Further research into how the housing market works, the accuracy of marginal prices, and other issues which logically precede the identification problem should be pursued.

TABLE OF CONTENTS

FIGURES

TABLES

CHAPTER 1

INTRODUCTION[1]

1.1 Benefit Cost Analysis and the Hedonic Model

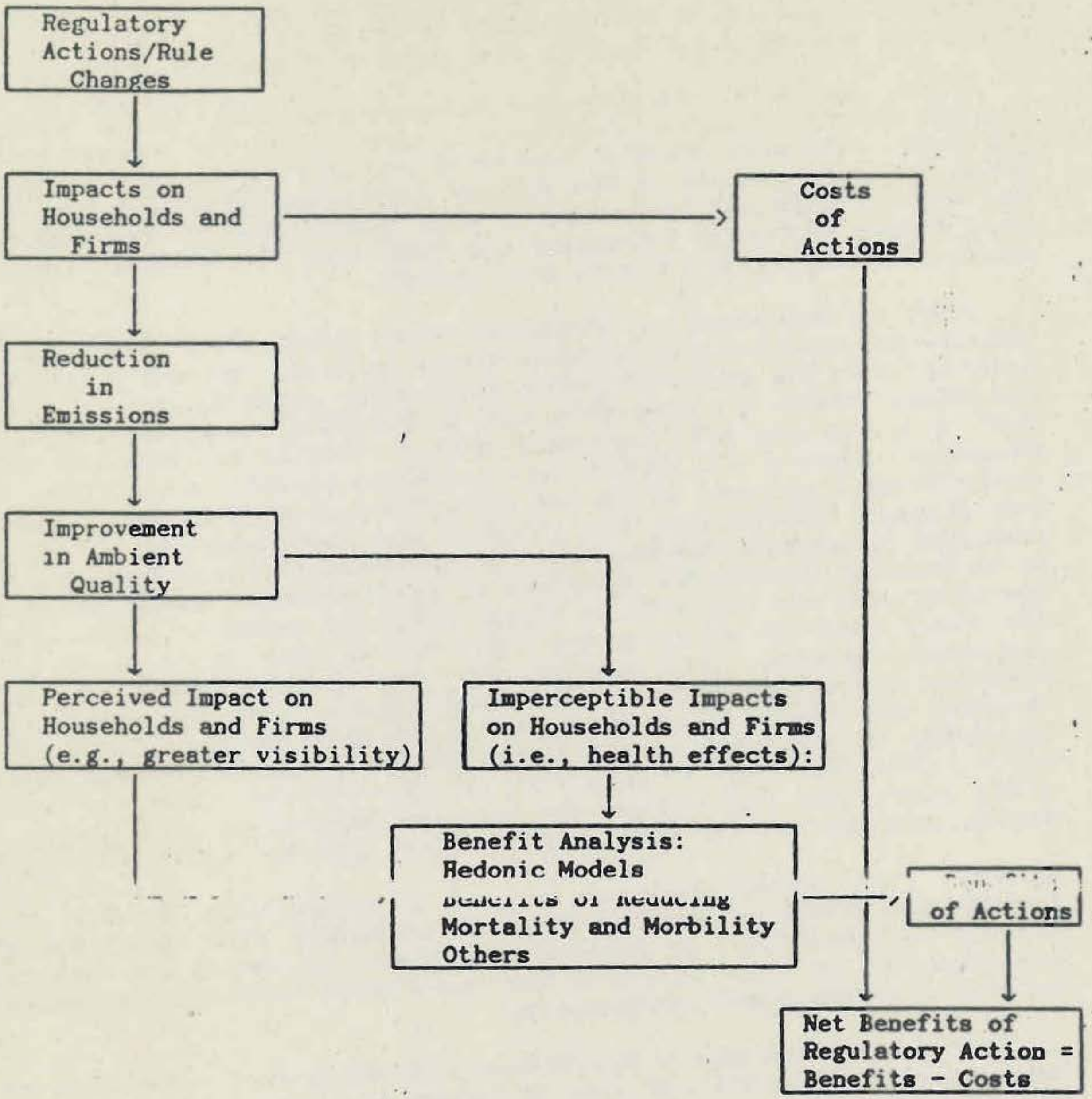
This report deals with one approach to inferring the value of environmental improvements--the hedonic method. It is part of the accepted wisdom of economics that environmental quality is a public good. Hence improvements in environmental quality will tend to be provided in less than optimal quantities by decentralized decisions. A corollary to this tenet is that government intervention may be required to provide optimal quantities of environmental improvements. To determine optimal quantities, the costs and benefits of environmental improvements are needed. In practice, optimal quantities of environmental improvements are almost never directly sought. Instead, government intervention for environmental improvements comes in the form of new rules or changes in rules. Benefit cost analysis can be applied to changes in rules to determine whether they are in the right direction. If enough rule changes are evaluated, then optimal quantities of environmental improvements can be achieved indirectly.

The hedonic method is one of several widely used approaches to measure the benefits of environmental improvements. It relies on individual choices in markets when the quality of the environment is one dimension of the quality of the good for sale. The basic approach of the hedonic method is to infer willingness to pay for environmental quality from market prices reflecting quality differences. This method is typically practiced by gathering data on the sales of goods, for example housing, and then showing with statistical methods the relationship between sales price and all the characteristics of this good, including practical measures of the quality of the environment. This relationship is called the hedonic price equation and the specific effects of pollutants on the sales price, as shown by statistical methods, have provided an important link in determining the benefits of environmental improvements.

The role of benefit cost analysis in general and the hedonic method in [illegible] and net benefit changes in Figure (1.1) (adapted from Desvousges, Smith and McGivney, 1983, page 1.2). A rule change or regulatory action is designed to force households or firms to reduce emissions. In cases of any consequence, the reduction of emissions requires changes in behavior which are costly to households and firms. Hence the initial economic effect of rule changes is to impose costs on economic units. If the rule changes are effective at reducing emissions, then they will improve the ambient environmental quality. Improvements in environmental quality will be valued by society. Improvements in environmental quality which are perceived lead some households and firms to change their behavior. Implicit market methods of

FIGURE 1.1

THE LINKS BETWEEN REGULATORY ACTIONS AND
THE NET ECONOMIC BENEFITS OF ENVIRONMENTAL IMPROVEMENTS



benefit measurement, such as the hedonic method, attempt to measure the changes in benefits by recognizing that rational, consistent behavior reveals information about preferences. When we reveal information about choices involving environmental quality which are explicitly or implicitly costly, then under some circumstances we can infer what people will be willing to pay for changes in environmental quality. Consider air quality improvements. As households perceive different air quality in different locations, they will change their behavior in a directly economic way by bidding up the price of sites which have improved air. The role of hedonic analysis is to use such information on behavior to infer the willingness to pay for improvements in air quality. The purpose of this volume is to assess the potential of the hedonic method for measuring the benefits of changes in environmental quality.

There are both administrative and economic reasons for wanting to improve benefit estimation techniques in general and the hedonic method in particular. The administrative impetus is provided by Executive Order 12291, which requires agencies of the Federal government to estimate the benefits and costs of major regulatory actions (with impacts greater than $100 million). Good benefit estimation techniques can help make the EO12291 a productive order. Bad techniques will make it a charade.

While the administrative procedures under which the Federal government operates are important and certainly should influence research in benefit-cost methods, there are additional cogent reasons for improving benefit estimation techniques. There is a compelling logic to benefit-cost analysis. Whatever its fault, it is the only fully consistent method available for assessing resource allocation. Hence it will tend to have influence, implicitly or explicitly, in the public decision process. In the use of benefit-cost analysis for environmental rule changes, benefits seem less plausible than costs because they come from intangible or aesthetic services that are not traded on the market. Costs tend to be incurred directly for purchases of physical capital goods or as higher operating costs and indirectly as higher prices for consumer goods. Further, the direct costs of environmental improvements tend to be borne by well represented groups. For example, air quality improvements may require expensive alterations of fossil fuel power plants. For any region we can describe the impact of rules about the sulphur content of coal or the installation of scrubbers on the stacks of power plants. We can also rest assured that the costs of such rule changes will find their way into the public debate over rule changes for they are incurred by small groups. But benefit estimates are far harder to introduce into the debate because they are harder to defend. The benefit estimates are at a disadvantage because of the metaphysical nature of benefits and the difficulties with techniques which estimate such benefits. So from the perspective of making the best use of our resources, we would do well to learn more about methods of estimating the benefits of environmental improvements.

The logic of economics in benefit-cost analysis is clear. Computing money measures of the benefits and costs of regulatory changes provides a common unit of analysis, and under the right circumstances, enables researchers to suggest when changes in rules are socially worthwhile. Yet, as Figure 1.1

shows, there is more to benefit cost analysis than simply measuring benefits. To determine economic benefits, the impact of rule changes must be traced through a variety of environmental and technical relationships. Further, as study of environmental decisions shows, there is more to the decision process in evaluating rule changes than the simple logic of calculating benefits and costs. These changes in economic welfare play a role in the decision process but so does information about who gets the benefits, and who incurs the costs, information about the effects of rule changes on emissions, emissions on ambient quality, and ambient quality on humans. Descriptive information about all the links in Figure 1.1 improves the cogency of analysis in part by reducing apparent uncertainty. Further, not all benefits and costs of equal magnitude carry equal weight in the decision process. It is the whole picture, from rule change to net benefit-cost analysis, including all the intermediate links, which determines whether proposed rule changes are enacted. Those analyses which appear more certain and which tell a more plausible story will be more convincing. Studies which communicate their results to a broader audience will be more effective, as will studies which provide a richer picture of the course of events.

What are the implications of such a pluralistic decision process for research on methods of benefit estimation? Should we abandon the attempt to develop logically consistent and plausible models of economic behavior for benefit measurement? We believe not, for two reasons. First, models which are logically consistent must help explain how people respond to changes in external circumstances, including changes in the economic rules of the game and changes in the natural environment. Such responses play a critical role in the link between rule changes and net benefits in Figure 1.1. Thus the effort to explain behavior in a consistent and plausible way, which is the essence of economic models, will help establish the framework not only for calculating benefits but also for describing the environmental links. Second, while benefit analysis works within the limited truth of logically consistent behavior, it is nevertheless our only tool for thinking systematically about scarce resources, whether environmental or other.

When we take a broad view of assessing the worth of rule changes, the hedonic method shows especial promise. At best, this approach would allow researchers to infer the value of changes in environmental amenities which result from the workings of a market. The potential advantage of this method over other methods, such as travel cost models or contingent valuation[2], is the presence of market prices which reflect differences in environmental amenities. At worst, the hedonic method [illegible]

environmental changes influence behavior, and adverse changes may make people worse off. Such scientific evidence can help establish the intermediate links in Figure 1.1. Evidence that environmental changes influence behavior is perhaps the weakest link in Figure 1.1, as we can learn from the General Accounting Office (1984) and Freeman (1982). Epidemiological studies do not always provide unambiguous evidence that air pollution affects human health. The adverse effect of water pollution on recreational activity is easy to imagine but there is little hard scientific evidence to document it. Thus part of the attraction of the hedonic method is its direct use of evidence. It shows in a way that noneconomists can appreciate how pollution affects well-being. If researchers can find a way to make the method yield measures of willingness to pay for changes in air quality, they will have an exceptionally valuable tool. If all we can salvage is evidence that air pollution affects housing values, we at least have evidence that pollution matters, which is often more than can be said now.

In the right circumstances, the hedonic method can be used to determine benefits of changes in public rules. There are several unsolved practical and conceptual problems involving the use of the hedonic models. The purpose of this report is to investigate the conceptual and practical problems of using hedonic models. The impetus for the research in this volume comes from the so-called identification problem in hedonic models. Solving the identification problem means developing the hedonic method so that it will tell us something about the preferences of individuals for environmental quality, and how individuals respond to changes in environmental quality. Without such information, the hedonic method can tell us only what emerges in the market, which reflects only one piece of information about preferences, the value of quite small environmental changes. Solving the identification problem means pushing the hedonic method to tell us more about the preferences of individuals behind the market, so that we know how to value large changes in environmental quality.

1.2 Overview of the Volume

The chapters in this volume are prepared by different authors or combinations of authors. While they all contribute toward the goal of the research, they may nevertheless be read independently of one another. Chapter 2 gives an assessment of the hedonic method as it is currently practiced, discussing the variety of its applications as well as its unsolved problems. Chapter 3 reviews current solutions to the identification problems and offers an interpretation of identification in a single market setting. Chapter 4 develops the structural system of which the hedonic equation is one part, and states the conditions for identification in a traditional econometric setting. Chapter 5 provides some evidence on estimation of the hedonic price equation in the form of Monte Carlo results. Chapter 6 creates a model which simulates the workings of a housing market and explores welfare measurement and choice of functional form in the hedonic price equation. Chapter 7 deals with the question of whether the hedonic model is appropriate for housing choices, and proposes several alternatives to current practices.

[illegible] Chapter 2 through 7 are rather diverse. Chapter 8, the conclusion, attempts to distill what has been written in the previous chapters as well as what has been learned on the project to provide an understanding of how to make the best use of hedonic models for measuring the benefits of environmental improvements.

1.3 Some Conclusions

The identification problem cannot be solved through empirical research. The identification problem deals with how much prior information one needs to bring to empirical analysis in order to recover the parameters related to preferences for environmental quality. In the hedonic case, we are concerned with the amount of prior information needed to identify the parameters of the preference function. Thus it is in the nature of our charge from EPA that our results are conceptual, not empirical. Empirical support, where provided, comes in the form of Monte Carlo or simulated markets, which allows the use of prior information.

Our findings with regard to identification are positive although heavily qualified. Chapters 3 and 4 demonstrate that identification of the preference parameters from single market data is possible, but only through the choice of functional form which is largely untestable. Chapter 5 is concerned with consistency in the estimation of the parameters of the hedonic price equation.

Our findings concerning the applicability of the Rosen version of the hedonic model are negative. Chapter 7 shows that the hedonic model is not well suited for locational choice. Chapter 6 demonstrates that applying different benefit measures from the Rosen model to changes in locational attributes can lead to vastly different results, a consequence of the disparity between choice in the hedonic model and locational choice. These conclusions relate to the use of hedonic models for valuing locational amenities, but not necessarily other uses of the hedonic model. Even when the hedonic model is not used for valuing locational amenities, one must still deal with the identification problems.

These conclusions suggest that environmental research which attempts to impute the benefit of improvements in air quality from the relationship between property values and air pollution should pursue new methods. In particular, methods which characterize the process of bidding for discrete bundles of attributes under uncertainty may prove fruitful.

CHAPTER 1

FOOTNOTES

1 Chapters with no authors listed (1, 2 and 8 and appendixes) were written by K. E. McConnell.

2 The travel cost method is an approach for evaluating recreation resources. It is useful also for valuing environmental amenities when they influence the quality of recreation. The method works by observing how people change their visits to a site as their costs increase. The contingent valuation approach works by asking an individual how much he would pay for hypothetical changes in environmental amenities. A thorough discussion of each can be found in Freeman (1979a).

CHAPTER 2

HEDONIC MODELS: CURRENT RESEARCH ISSUES

2.1 Introduction

The purpose of this chapter is to provide a brief introduction to hedonic models and to outline the chief research issues currently facing practitioners. The chapter will not attempt a survey of the literature, nor an exhaustive catalogue of issues raised by the hedonic method. The emphasis here will be on the use of hedonic models for measuring benefits of environmental improvements, especially through the relationship between housing values and air quality.

2.2 Choice of Quality and the Hedonic Model

This research investigates the hedonic method, yet this method encompasses a fairly broad range of approaches. In practice, the term hedonic has come to mean any method valuing the quality of a good through measuring its demand. In the context of environmental research, hedonic tends to mean any method which values the public good — environmental quality -- through information on purchases of a private good. Our focus will be narrower, specifically on the Rosen model, but it will be useful to survey briefly the origin of various approaches which go by the name of hedonic.

Models of quality may be examined along several different lines. For example Hanemann (1981) distinguishes between the "differentiated" and "generalized" approaches to demand analysis, depending on whether goods with different quality characteristics are treated as separate commodities or the same generalized commodity. In the current discussion, we will consider two types of quality models: those in which the consumer chooses quality in a vector of n dimensions and those for which quality may be measured as a scalar. While the distinction may occasionally appeared blurred on close examination, it will serve our purpose for the analysis to follow.

[illegible]

originated with the work of Houthakker (1952) and Theil (1952), though Houthakker only analyzed the case where each commodity has only one dimension of quality. (Houthakker cites the prior work of Court (1941)). Work by Adelman and Griliches (1961) is a direct descendent of the Houthakker work and provides the initial theoretical basis for the use of hedonic price indexes. Adelman and Griliches posit a preference function of the form

$$U = U(x_1,...,x_m,z^1,...z^m)$$

where x_i (i=1,m) are commodities purchased on the market and z^i (i=1,m) are n_i dimensional vectors measuring the attributes of commodity i. All elements of the preference function are subject to choice, and the price of the i^{th} commodity is also a function of its vector of characteristics:

$$p_i = p_i(z^i).$$

The hedonic method as an index number practice was originally applied to automobiles by Griliches (1961). Additional applications may be found in Griliches (1971). Work by Becker (1965) and Lancaster (1966) is similar in the sense that it involves quality choice in a large number of dimensions, but does not directly tie into the hedonic practices.

The hedonic models differ from the Becker-Lancaster models of household-produced commodities by having a market interposed between household choice and prices reflecting quality. This market was typically assumed to exist, in the sense that prices reflect quality but there was no formal demonstration of why market prices reflect quality. This gap was filled by Rosen (1974) who showed how buyers and sellers of a good with measurable attributes establish a price locus reflecting those attributes. This locus can be taken as a given by any single buyer, who then chooses the kind of good to buy by choosing the optimal quantity of each attribute.

The choice along one dimension, or the exogenous scalar influencing the quality of a private good, represents the alternative modeling approach. This approach seems to have been developed independently by several different people. Maler (1971, 1974) developed the theoretical conditions for measuring the value of a public good by examining purchases of private goods. Quantities of the public good influence the quality of the private good, as for example, water pollution might measure the quality of recreation trips. Stevens (1966) provided an application, without the theoretical qualification. Bradford and Hildebrant (1977) provide theoretical results similar to Maler. These results are extended by Willig (1978). Fisher and Shell (1968) developed a model which is also relevant because, while they were interested in price indices, they limited their analysis to one dimension.[1]

The distinction between the number of dimensions is especially crucial when we consider the location decision. By its nature it is limited to two dimensions, and typically converted to one dimension, the distance from the center of the city. Thus, for example, the location model of Alonso (1964) is similar to the single public good case of Maler, Bradford and Hildebrandt and Willig.

Models for estimating the effect of the quality of a commodity cover a broad spectrum. These models, have all come under the rubric "hedonic", broadly interpreted. We are interested in a narrow segment of hedonic models, the Rosen model. In the following section we discuss its use in environmental economics.

2.3 The Hedonic Model in Environmental Economics

In concept, hedonic models provide information on the willingness to pay for public goods because preferences revealed for private goods in part reflect the demand for public goods. Private goods which provide better access to public goods, such as cleaner air or more quiet, will be valued more highly by households, and private transactions will reflect the value of public goods. The hedonic model is both a theory and an empirical method which attempts to separate the effect of qualities such as access to public goods from other influences on the price of private goods. Like several methods for assessing the benefits of environmental improvements, the hedonic method of valuing the environment began as an empirical approach. Ridker and Henning (1967), Nourse (1967), and Anderson and Crocker (1971) analyzed the effect of air pollution on housing values. Their empirical results and analytical efforts to understand their empirical results spawned a lengthy debate over the method. The development of the Rosen model played an important role in settling some of the issues debated.

The initial applications of the hedonic method to environmental quality attempted to infer willingness to pay for changes in air quality from housing prices. In current environmental work, applications of the hedonic method to the air quality-housing price case predominate. However, the first application of hedonic models was to automobiles, with subsequent applications of the hedonic method to labor services (hedonic wages), and other goods and services.

The promise of the hedonic method can be gauged by the number and variety of applications in the current literature. Under the rubric of air pollution, a number of different pollutants have been valued. For example, Palmquist (1983a) investigates the effect of total suspended particulates, nitrogen dioxides, sulphur dioxide and ozone on property values in 14 cities. Bender et al. (1980), Li and Brown (1982), Schulze et al. (1983), and Harrison and Rubinfeld (1978) (among many others) have also estimated the relationships between housing prices and air pollution.[a] In addition, other environmental effects have been measured using the hedonic model. Noise (Nelson, 1978; Li and Brown, 1982), accessibility to shore line, (Brown and Pollakowski, 1977; Milon et al., 1983) and water pollution (Epp and El-Ani, 1979; Rich and Moffit, 1982) have all been shown to influence housing prices. Work to determine the effect of proximity to hazardous waste sites on housing values is also proposed or under way. The hedonic model has been used or [illegible] which account for the attraction of the house, for example, schools and crime (Jud and Watts, 1981; Bartik and Smith, 1984), threat of earthquake (Brookshire et al., 1984), climate (Freeman, 1984) and many kinds of urban amenities (Bartik and Smith, 1984). Of course, all aspects of the house itself have shown to be influential in determining housing prices, for example, size and number of rooms, presence of air conditioning, swimming pool, fireplace, detached garage, number of bathrooms, age, type of construction, etc. (Palmquist, 1983b; Li and Brown, 1982).

The consistency of findings, especially with regard to air pollution, has been as impressive as the variety of applications. While several papers skeptical of the relationship between housing prices and pollution appeared in the 1970's (Wieand, 1971; Smith and Deyak, 1975), recent work has supported the relationship. Published research tends to show that higher levels of air pollution are correlated with lower housing prices, cet. par., though it may be that positive or inconclusive findings are less likely to get published. Somewhat more surprising is the result from hedonic wage models that wage premia are associated with higher air pollution (Bayless, 1983; V. K. Smith, 1983). Thus the hedonic models show their promise through the variety of applications and the consistency of findings. Perhaps most important, the basic model is intuitive and easy to explain to noneconomists.

The two types of models discussed in the previous section are useful for examining some work which occassionally goes under the rubric hedonic. Polinsky and Shavell (1976) and Polinsky and Rubinfeld (1977) have developed empirical models where the bid for each location depends on the attributes of that location. These models involve optimization in one dimension, and hence are similar in spirit to the second category of models, the single public good of Maler, Bradford and Hildebrandt, and Willig. Thus, the work of Polinsky and Shavell and Polinsky and Rubinfeld may be considered hedonic, but because it involves only one dimension of choice, it is different from the Rosen model.

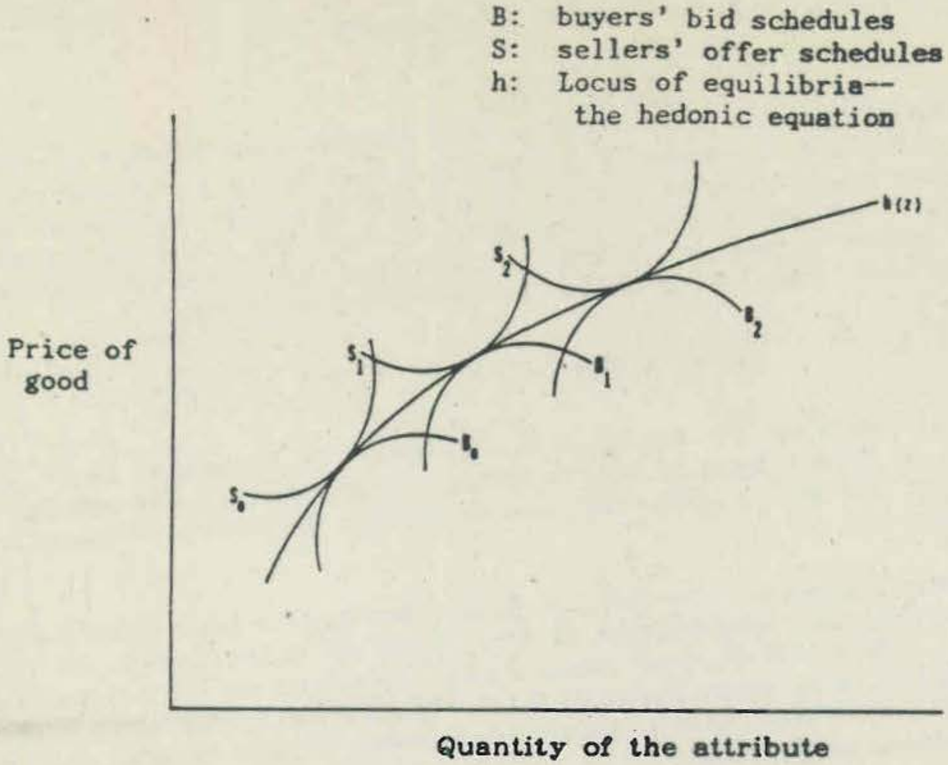
2.4 The Basic Rosen Model

Despite the promise of the hedonic method, there remains a number of problems which arise in its application. Before spelling out the nature of these problems, it will be useful to give some structure to the Rosen version of the hedonic method. The following gives a skeletal version of the hedonic model, which was given its conceptual framework by Rosen.

Suppose that a market exists for a good with several attributes of quality. Wine may have sugar content, hue, and bouquet, or many more chemically measurable attributes. A house has windows, lot size, rooms, square feet, carports, etc. Cars have horsepower, length, acceleration. Sellers are aware of the costs of producing the good with different attributes. Buyers know that units of the good with different attributes bring different utility levels. When the market is relatively dense, that is, almost any level of attribute is technically feasible and may be supplied, and demanded, then we can equate [illegible]

Figure 2.1

Assume that there is only one attribute of the good, and it is measurable. Consumers come to the market willing to pay more for a unit with more of its attribute. This information is revealed by their bid functions, B_0, B_1, B_2, which differ if they have different preference functions or different incomes. Sellers know the extra cost of producing the good with more of the attribute, and because there are sellers with different characteristics, they offer different quantities of the attribute at different prices, denoted by the schedules S_0, S_1, S_2. The market equilibrium yields the hedonic price equation denoted



The Basic Hedonic Model
Figure 2.1

h, which is a locus of equilibrium points of various quantities of the attribute. Individual buyers or sellers take the hedonic price relationship as given and make their marginal selling or buying decisions according to its implicit trade-offs. Buyers choose goods which equate the marginal value of the attribute with its marginal cost, given by the hedonic equation. Sellers produce goods which equate the marginal cost of production with the marginal returns, also given by the hedonic equation. This model will be the source of much greater scrutiny late in this volume.

The structure of the model given above was developed persuasively by Rosen. The estimation methods were also codified by Rosen in the following [illegible] attributes. For example, housing price depends on the site-specific attributes, neighborhood characteristics and environmental quality. The resultant relationship is the hedonic price equation. Second, compute the partial derivative of the hedonic price with respect to the i^{th} attribute, and use this as an endogenous marginal price in a model of supply and/or demand.

It will aid our discussion of the hedonic method to be more specific about the two step approach. Let us assume that we analyze buyers' choices, and hence are interested in parameters of preferences. Let

$$p = h(z;\gamma) \tag{2.1}$$

be the hedonic price equation, where z is a K-dimensional vector of attributes of the good and γ is a vector of parameters describing the function. Using best fit methods, we estimate (2.1). In equilibrium, the consumers' marginal bid for the attribute will equal the marginal cost of the attribute, as given by the hedonic price equation. Then we use the predicted derivative as a dependent variable, marginal price, in the following equations:

$$\partial h/\partial z_i = m_i(z,y;\beta) \qquad i = 1,\ldots,K \tag{2.2}$$

where m_i is the marginal bid function (marginal to the functions B_0, B_1 and B_2 in Figure 2.1), y is income and β is a vector of parameters describing tastes. Expression (2.2) is the equilibrium condition for individual buyers in the hedonic market.

The economic framework created by Rosen has been rather widely accepted as providing a plausible explanation of the effect of amenities on the price of private goods. While there have been many questions about procedures for applications, there have been few about the theoretical structure. Especially in the areas of urban and environmental economics, it has become part of the accepted theoretical structure.

2.5 Some Research Issues

While the Rosen model of hedonic pricing has served well in its positive role, questions arise when we try to use the model for normative purposes. For example, the hedonic equation may do well in predicting the cet. par. effect of another bathroom on the price of a house, but it is less clear what it reveals about the welfare effects of a decrease in total suspended particulates. Further, there are some ambiguities about the applicability of the Rosen model to the choice of housing location. In this section we survey several questions currently debated in the literature. These questions are important because they relate to the use of the hedonic method for measuring the changes in environmental amenities, but they in no way exhaust current research topics. A discussion of these issues will help in understanding the focus of this volume.

We can illustrate current research directions by referring to the basic model and to expressions (2.1) and (2.2) and to figures similar to Figure 2.1. We divide the research topics into five areas:

1. What practical problems arise in estimating the hedonic price equations?
2. Can the parameters (β) of the m function in equation (2.2) (typically called the inverse demand function or marginal bid function) be identified, and if so are there serious estimation problems which then arise?

3. How can the welfare changes induced by exogenous changes in attributes be measured?
4. Does the hedonic model capture all the welfare change associated with changes in an environmental attribute?
5. Are the structure and assumptions underlying the hedonic model appropriate for the issues relating to choice of location by households? That is, when households choose the location of their residence, is the hedonic model working?

Considerable effort has been directed to problems encountered in estimating the hedonic price equation, the first topic. Four of the problems that arise in fitting the hedonic price equation are multicollinearity, selection of functional form, measurement of the amenities or attributes, and the aggregation issue. The collinearity problem is especially severe. Bigger houses typically have more of all kinds of attributes - bath rooms, lot size, garage space, and a higher likelihood of having amenities which come in discrete units - pool, air conditioning, a scenic view. Amenities within a community tend to be highly correlated. Localities with good schools tend to have nice park systems as well as high tax rates. Different air pollutants are particularly likely to be correlated. Weather patterns and location close to common emission sources cause some areas to have more of all pollutants than other areas. Collinearity is probably most severe for the characteristics specific to the house. It would be wrong, however, to argue that multicollinearity is always a problem. Palmquist (1983a) has shown that for one set of 14 cities, collinearity is not a problem for pollutants.

The choice of functional form for the hedonic price equations is a critical one, in that it determines how marginal prices behave. Yet by the nature of the model, we can expect little or no theoretical guidance for choosing among alternative functional forms. As can be seen from Figure 2.1, the hedonic equation is a locus of equilibria, and has embodied in it the structural aspects of buyers and sellers. Best fit methods, such as Box-Cox approaches used by Bender, Gronberg and Hwang (1980), Halvorson and Pollakowski (1981) and others seem appropriate, but these methods may not result in well-defined maxima for households with quasi-concave preference functions. Closely related to the choice of functional form is the problem of complete specification. It is virtually impossible to specify a hedonic equation which includes all the attributes which influence price. The exclusion of collinear attributes can have two affects. First, it can bias the coefficients of the hedonic equation. Second, when combined with nonlinearity, such misspecification creates errors in the hedonic price equation. Because of the nonlinearity these errors are transmitted to the estimated marginal price, and are quite likely to be correlated with any instruments (such as income) used in the estimation of demand relations. (See Epple, 1982, and Bartik, 1983).

The measurement of pollution variables is an important issue. The theory requires that all market participants respond to the same attributes, but perception of air quality may vary substantially across households. And perceptions may not be closely linked with actual measures of pollutants, available for example, from monitors. The problems of multicollinearity and amenity measurement complicate one another, because it is doubtful that a single air pollutant can capture households' perceptions of air quality. Yet if the pollutants are highly correlated, it will be quite difficult to separate their effects. The work by Palmquist (1983a) on creating an air pollution index is quite promising in this regard, because it is a first attempt to compute an index which might replicate households' perceptions. Further, Palmquist has shown for at least one set of 14 cities that collinearity is not especially severe for the pollutants. Bartik and Smith (1984) have highlighted the problem of perceptions.

Finally, there is the question of aggregation of observations. Early work such as that by Ridker and Henning used median sales price of owner-occupied housing, where the census tract was the unit of observation. But recent empirical research has relied predominantly on housing sales data or homeowner opinion surveys. The question of when and whether parameters of the hedonic equation can be recovered from aggregate data has received no formal attention.

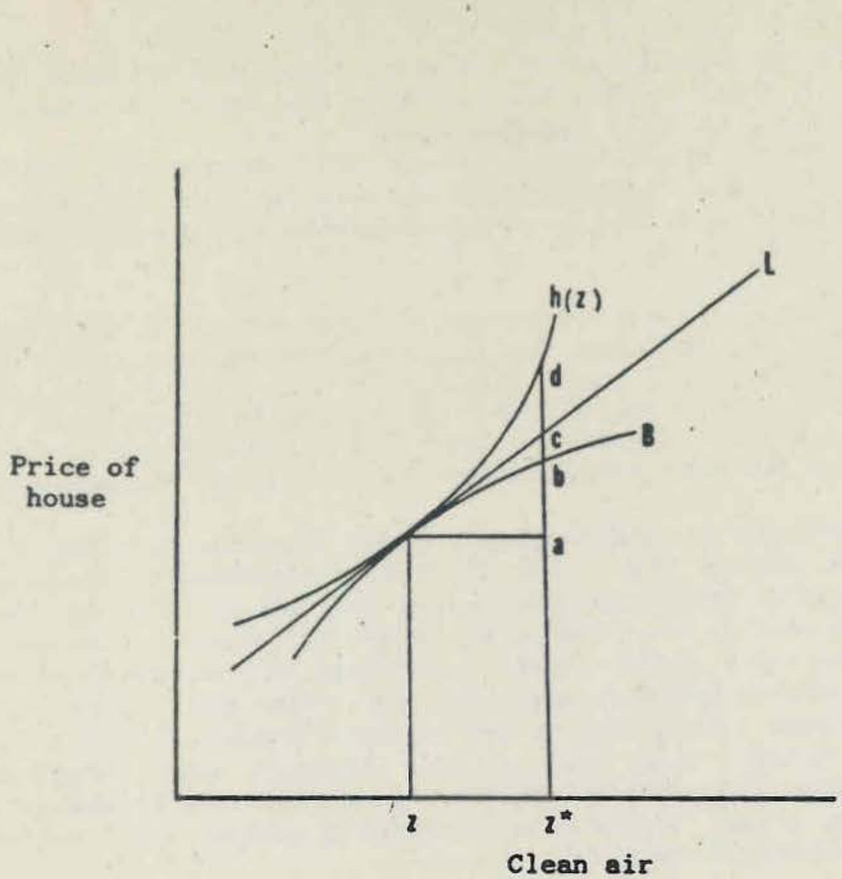
The second issue in the research list is the identification problem. The nature of this problem can be understood by rewriting equations (2.1) and (2.2) as

$$p = h(z;\gamma) \tag{2.3}$$

$$\partial h(z_j,\gamma)/\partial z_i = m_i(z,y;\beta) \quad i = 1,\ldots,K \tag{2.4}$$

where γ is a vector of parameters describing the hedonic price equation and β is a vector of parameters describing the marginal bid function. The Rosen two step approach estimates (2.3) first, and then uses the predicted derivative to estimate (2.4). The identification problem in an intuitive sense comes from having estimates of β actually be combinations of γ and β. Brown and Rosen (1982) give the best illustration of this particular problem. The issue is currently receiving as much attention as any other issue in hedonic models. As we show next, the identification problem is important to the extent that information about preferences for environmental amenities is needed. It is possible, however, that benefit measures can be computed without such information.

The third issue deals with the way welfare measure can be derived from the hedonic method. There has been surprisingly little research on this topic, especially since for environmental matters, welfare analysis plays such a central role. The basic issue can be stated as follows: Suppose that government actions can cause the attributes of housing to be improved by reducing air pollution. How should the hedonic model be used to measure the economic benefits of better air? The problems surrounding this issue can be addressed with Figure 2.2. This figure shows the bid function (B) for an individual, and the market hedonic price function, h(z). Suppose a government rule results in an increase in the amenity — better air — which is experienced by the individual as an increase from z to z*. Assuming the individual to be in equilibrium at z, we can discuss three measures of welfare changes commonly used in the literature:



Welfare Measures for Increasing an Attribute

Figure 2.2

(i) the household's increase in willingness to pay for the site: ab

(ii) the predicted increase in the price of the site, based on the hedonic price equation p(z): ad

(iii) a first order Taylor's series approximation of (i) and (ii). L is tangent to the equilibrium at e, so that its slope is equal to the common marginal price - marginal willingness to pay, and an estimate of (i) or (ii) based on a linear extrapolation is given by ac.

Δ hedonic price $>$

linear extrapolation of Δ hedonic price $>$

Δ willingness to pay

The consumer's willingness to pay is a superior measure, but requires knowledge of the parameters β of m in (2.4) and requires successful completion of the Rosen two-step approach. The linear expansion of B or h is most often used and most criticized. Its accuracy can be seriously impaired by two possibilities:

a) Only a few combinations of the z's are available in practice so that there is no equilibrium of marginal price and marginal willingness to pay. In fact, unless the available z's are quite dense, a negative marginal bid is quite possible for z.
b) The hedonic price function need not be convex; for equilibrium purposes it need only be less concave than the bid surface. In that case, linear extrapolation of p will exceed the prediction made by $p(z^*) - p(z)$ given in (ii).

Thus it seems that benefit measures will be improved by recovering the parameters of the bid function, but this requires solution of the identification problem. If the hedonic price equation is not "too" convex, then it may provide a decent estimate of the value of changes in z. At least we know that whether we use the prediction from the hedonic price or its linear extrapolation, we will have overestimated the change in willingness to pay.

Another difficulty in welfare measurement becomes apparent when we look more closely at Figure 2.2. The household equilibrium requires tangency between the hedonic price equation h and the bid function B. The tangency exists at z, but not at z*. Hence the measures described above are, in the phrase of Bartik and Smith (1984), restricted partial equilibrium measures. They are restricted because they do not allow the market to adjust to changing conditions. When the z's are changed exogenously, the initial supply conditions no longer hold, and a new hedonic price equation must be established. The appropriate welfare measures require comparing an old equilibrium with a new equilibrium, something which the "restricted partial equilibrium" measures do not do.

The fourth topic given above also involves welfare measurement. The essence of this problem concerns potential double counting of benefits from an environmental improvement. To what extent does the hedonic method applied to property values measure benefits that might also be captured by other methods? Roback (1982) has investigated the case when wages are influenced by environmental attributes. Other cases remain to be investigated. The economic use of epidemiological studies attempts to measure the benefits of improving air quality, which is also the role of housing value studies. Location near a clean water site may be capitalized into land prices, and hence measure in part the demand for travel to the clean water. A separate but related issue concerns the purchase of attributes which reduce the effect of pollution, for example, air conditioning. Because people spend a majority of time indoors [illegible] expenditures [illegible] devices [illegible] air conditioning can avert some effects of air pollution. These issues must be worked out in concept before we can investigate their practical importance.

Research on the fifth topic has addressed two questions, both especially problematic for the real estate market. The hedonic model assumes that the goods are sold at auction with buyers and sellers having full information. The housing market, in fact, is one of sequential bids and substantial uncertainty about the hedonic locus. Work by Ellickson (1981), Lerman and Kern (1983) and Horowitz (1983) is designed to model the housing market to reflect more accurately the way transactions are made. Another important assumption in the hedonic model is the continuity of the hedonic price function in attributes of the good. Continuity assumptions are routinely made and violated in economics, usually with little impairment of conceptual or empirical analysis. Continuity assumptions may not be so innocuous in hedonic models. Housing attributes such as rooms, air conditioning, and swimming pools not only are not continuous but are typically available in only a few combinations. Further, because of the limited number of bundles available, choices may tend not to equate marginal bids with marginal costs. The lumpy aspect of housing, implying discrete choices, is modelled initially by McFadden (1978). This particular aspect of hedonic models is a fruitful area for research.

For purposes of this volume, the issues raised above fall into two categories. On the one hand there are the important practical problems involving estimation of the hedonic equation, determining what benefits can be calculated from the hedonic model, and the accuracy of various restricted measures of welfare changes. These problems are not different from the problems one confronts in any kind of empirical work in economics. They are primarily the consequence of less than perfect data. On the other hand, there are the issues of identification and whether the hedonic model is appropriate for the choice of residential location. These issues have the common aspect that their solution does not hinge on better data. The problem of identifying parameters of preference functions when households have nonlinear budgets is severe even with perfect data. Further, if the hedonic model is not the right model for choice of location of residence in concept, no amount of data will make it so in practice. This volume is concerned with problems of the second sort. That is, we will investigate those issues which in principle may prevent the method from providing useful input to benefit-cost analysis.

2.6 The Charge of the Research

This research was undertaken as a part of research project on implicit market methods of measuring the benefits of environmental changes. The explicit charge for the hedonic research is to "develop solutions for the underidentification of hedonic demand curves for environmental public goods and demonstrate, using suitable pollution problems" (EPA Request for Proposal, April 1983).

This charge has been the driving force of our research. But we have expanded our research to those topics which in principle prevent the

interpreted the identification problem here as the problem of recovering the parameters of a function which yields willingness to pay by households for changes in attributes of a good. That is, we wish to ascertain under what circumstances we can recover the parameters of the $m(z,y;\beta)$ function in equation (2.4) because recovering these parameters may help improve welfare measurement. The following two chapters explore directly the identification problem. These chapters are quite different in approach but have in common the idea that identification is solved in concept. Other chapters, too, are concerned with whether the hedonic method works in concept.

CHAPTER 2

FOOTNOTES

1 Muellbauer (1974) indicates how the distinction between types of models can become blurred. He increases the quality dimension of the Fisher-Shell model to make it a choice of several attributes and reduces the dimension of the Houthakker model to make it a one dimensional choice. As we shall argue later, the choice of model ultimately depends on the technical and institutional characteristics of the problem.

2 For further works on pollution and property values in the hedonic model, see the references at the end of this volume, Bartik and Smith's (1984) references, and those provided by Rowe and Chestnut (1982).

CHAPTER 3

IDENTIFICATION OF HEDONIC MODELS

Robert Mendelsohn[1]

3.1 Introduction

Although the theory and econometrics for understanding markets for homogeneous goods have been understood for decades, the problems of modelling markets for heterogeneous goods has received attention only recently. One fruitful approach to dealing with heterogeneous goods has been the hedonic model. The heterogeneous good is envisaged as a bundle of homogeneous attributes. For example, a residence is composed of attributes such as the number of rooms, lot size, school quality, air quality, and other characteristics. From the work of Court (1941) and Griliches (1971), it is now commonplace to estimate the implicit prices of these attributes by regressing expenditures on the bundle (the price of the heterogeneous good) upon the observed attributes. As noted by Rosen (1974), the resulting marginal price gradient is the locus of market prices which equilibrate demand and supply. For marginal valuations, this locus is all that is needed. However, for nonmarginal valuations where the observed price gradient is expected to change in response to some policy of interest, it is necessary to uncover the underlying structural equations of the model. The purpose of this chapter is to discuss when and how the demand and supply curves for characteristics can be identified with available data and econometric techniques.

The first discussion of the identification problem with hedonic markets was raised by Rosen (1974) in his development of the basic hedonic market model. Rosen perceived the hedonic structural equations to be no different from traditional market models. He consequently asserted that the identification issue was just the familiar problem of sorting out supply from demand.

More formally, suppose the hedonic price function for the good is:

where z is a vector of attributes. Then the price gradient (of marginal prices) for each attribute z_i is:

$$p_i(z) = \frac{\partial h(z)}{\partial z_i} . \qquad (3.1)$$

The underlying inverse supply (g) and demand (f) functions for the attributes are:

$$p_i(z) = f(z,y) + \varepsilon_1$$

$$p_i(z) = g(z,w) + \varepsilon_2 \qquad (3.2)$$

where y and w are exogenous demand and supply shift variables, respectively, and ε_i are random error terms. Rosen recommended that the hedonic price function be estimated by OLS in a first step. Taking the derivative of the hedonic price function, the appropriate marginal price for the observed purchased bundle z would then be the dependent variable in the estimation of the structural equations (3.2). The identification problem, according to Rosen, is the separation of demand from supply effects.

Brown and Rosen (1982) offer an alternative identification problem. They are concerned about the use of the predicted marginal price from the price regression (3.1) in the structural equation estimation (3.2). They note that with linear functional forms, the variation in z captures all the variation in the predicted price. That is p_i is constructed:

$$\hat{p}_i = \hat{\gamma}_0 + \hat{\gamma}_1 z.$$

Thus, to estimate demand by regressing p on z and other shift variables y such as:

$$\hat{p}_i = \beta_0 + \beta_1 z + \beta_2 y$$

one should expect $\hat{\beta}_0 = \hat{\gamma}_0$, $\hat{\beta}_1 = \hat{\gamma}_1$, and $\hat{\beta}_2 = 0$ because there is no random variation in p that cannot be perfectly explained by z. Furthermore, at least with a linear marginal price model, the linear structural equation will always be the best fitting functional form because it provides a perfect fit. The structural estimation consequently just reproduces the original marginal price equation. The structural equations remain unidentified.

A third perspective is voiced by Mendelsohn (1980), Bartik (1983) and Diamond and Smith (1985). These authors note that maximization of profits or utility subject to the nonlinear budget constraint of a single price gradient results in only one observation for each actor in the market. Each of the [illegible] are substantively different. The identification problem in hedonic markets is not between demand and supply per se but rather between the response of one demander to one price versus a different demander to another price.

There are consequently three potential identification problems with single market hedonic models. (1) The "garden variety" simultaneity of demand and supply; (2) the use of estimated prices in structural equation estimation; and (3) the separation of price effects from shift effects across consumers or across suppliers. Corresponding to each of these problems, authors have recommended specific solutions.

In Section 3.2, we discuss solutions to the "garden variety" identification problem and demonstrate that traditional solutions are not adequate because of the simultaneity of shift and price effects. In Section 3.3, we review the use of predicted marginal prices in the structural equation and show that the Brown and Rosen critique can be generalized to any structural equation where the exogenous shift variables are additive. We further show that the estimation of prices is not the central problem. In Section 3.4, we address the special identification problem of hedonic markets, the untangling of price and shift effects. In this section, we show how nonlinearity in the price gradient and restrictions on the structural equations can lead to identification. The identification problem in the Brown and Rosen linear model can disappear in nonlinear models.

Finally, in Section 3.5 we discuss how observations from multiple markets (either intertemporal or cross sectional) can overcome the identification dilemma in certain circumstances.

3.2 Simultaneous Demand and Supply

If the inherent identification problem of hedonic models is the simultaneity of supply and demand equations, there are several plausible solutions. As recommended by Nelson (1978), Linneman (1981), and Rosen (1974), one could use econometric techniques such as instrumental variables or two stage least squares to separate demand from supply. For example, suppose the underlying model is:

$$z = f(p,y) + \varepsilon_1 \tag{3.3a}$$

$$z = g(p,w) + \varepsilon_2 \tag{3.3b}$$

where $f(\cdot)$ is demand, $g(\cdot)$ is supply, y and w are shift variables, and ε_1 and ε_2 are error terms. Marginal price p is endogenous in this model, being the result of both supply and demand factors. Consequently p is affected by both ε_1 and ε_2 and so is correlated with both. OLS regressions with p would be biased. To correct this problem, one regresses p on the exogenous shift variables y and w. The resulting predicted level of price, p, can then be entered into either structural equation for second stage estimation.

An alternative way to control for the simultaneity of supply and demand is to assume one of these structural equations is fixed. For example, Harrison and Rubinfeld (1978) assume that the supply of clean air is unresponsive to the price of clean air. As Nelson (1978) and Freeman (1979a) note, the level of air quality in each area may indeed be insensitive to the prices charged in each housing market. However, the supply of clean air is the amount of housing available with clean air, not the amount of acreage available. Consequently, builders could provide more housing per acre in clean air locations if the price of clean air were sufficiently higher. Thus, it may often be inappropriate to assume that supply functions are perfectly inelastic in hedonic markets.

Parsons (1983) and Epple (1982) demonstrate that the identification problem in hedonic single markets deals with more than the traditional separation of demand and supply. Both these authors show that the traditional methods used to untangle demand from supply will not work in the single market context. Along a nonlinear price gradient, suppliers and demanders arrange themselves according to their underlying shift parameters y and w. This sorting procedure means that certain types of suppliers will tend to match up with particular demanders to transact special bundles along the gradient. For example, with housing, builders of homes in the outlying suburbs will tend to supply the attribute clean air. Demanders of clean air, possibly asthmatics, will tend to purchase these outlying homes. The introduction of the variable, asthmatics, will represent the builders of outlying homes just as much as the demanders for these clean air homes. The single market results in a one-to-one correspondence between particular demanders and suppliers, making it difficult to identify either structural equation. Thus, the identification problem is clearly more than "the garden variety" found in +traditional goods markets. The untangling of supply and demand is just at the surface of the problem.

3.3 Predicted Prices

Brown and Rosen (1982) show that when both the price gradient and the structural equations are linear, the predicted marginal prices cannot be used to identify the structural equations. The linear estimation of the structural equation simply reproduces the coefficients of the hedonic price gradient.

Brown and Rosen's proof can be generalized. Regardless of the shape of the price gradient, any structural equation which is additive in the exogenous shift effects will merely reproduce the price gradient. For example, suppose the price gradient is

$$p_i(z) = \gamma_0 + \gamma_1 z^3 + \gamma_2 \log z = \frac{\partial h(z)}{\partial z_i} \tag{3.4}$$

Any structural equation which additively includes $p_i(z)$ will reproduce (3.4). For example:

$$p_i = \beta_0 + \beta_1 z^3 + \beta_2 \log z + \beta_3 z^\alpha + \beta_4 y$$

$$p_i = p_i(z) + 0\, z^\alpha + 0\, y.$$

That is, the estimated coefficients on z^α and y would be zero.

To surmount this problem, analysts have restricted the family of structural equations so that none of the members can have the above properties. Brown (1983) suggests omitting particular expressions for z in the structural equation which are in the hedonic equation. For example, one could leave out the log z term found in (3.4). Alternatively, one could omit a particular attribute z_k in the structural equation. Finally, one could adopt a different functional form (log linear, linear, or semilog) in the hedonic price versus structural equations. This latter approach is used by Harrison and Rubinfeld (1978), Nelson (1978), Linneman (1981), Witte et al. (1979) and Bloomquist and Worley (1981) in their hedonic models.

Although the alteration of functional form between the hedonic and structural equation leads to different parameters between the two equations, it is not clear whether the assumption has identified the true underlying structural equations. After all, making different assumptions about the shape of any of the curves leads to different parameters. Although the Brown and Rosen (1982) model has touched the surface of an identification problem, the paper provides little guidance to the underlying cause of the problem or to its appropriate solution.

In order to show that the problem with hedonic markets is not the use of estimated prices, let us reproduce the Brown and Rosen model and show that the structural equations are not identified even when the price gradient is observed (not estimated). To keep the notation simple, let us assume the marginal price is a linear function of a single attribute:

$$p_i(z) \equiv \frac{\partial h(z)}{\partial z} = \gamma_0 + \gamma_1 z.$$

The structural equations are also assumed to be linear:

$$p_i = \beta_0 + \beta_1 z + \beta_2 y \qquad \text{(demand)}$$

$$p_i = G_0 + G_1 z + G_2 w \qquad \text{(supply)}.$$

Because the price gradient is the locus of equilibrium points between supply and demand, for each z, it must be true that:

$$\gamma_0 + \gamma_1 z = \beta_0 + \beta_1 z + \beta_2 y = G_0 + G_1 z + G_2 w\ .$$

Solving for y and w respectively:

$$y = \frac{\gamma_0 - \beta_0}{\beta_2} + \frac{(\gamma_1 - \beta_1)z}{\beta_2} \qquad (3.5a)$$

$$w = \frac{\gamma_0 - G_0}{G_2} + \frac{(\gamma_1 - G_1)z}{G_2} \qquad (3.5b)$$

If we can observe z, y, and w for all pairs of demanders and suppliers, then we could estimate:

$$y = \hat{A}_0 + \hat{A}_1 z \tag{3.6a}$$

$$w = \hat{D}_0 + \hat{D}_1 z \tag{3.6b}$$

Suppose we also could observe the marginal prices so that we could know γ_0 and γ_1. The issue is whether the β and G parameters of supply and demand could be identified. If the problem is only with the use of estimated prices, the equations should be identified.

For the data to be consistent with both (3.5a) and (3.6a), it must be true that:

$$A_0 = \frac{\gamma_0 - \beta_0}{\beta_2} \qquad \text{(demand)}$$

and

$$A_1 = \frac{\gamma_1 - \beta_1}{\beta_2} .$$

Similarly using (3.5b) and (3.6b), it follows that

$$D_0 = \frac{\gamma_0 - G_0}{G_2} \qquad \text{(supply)}$$

$$D_1 = \frac{\gamma_1 - G_1}{G_2} .$$

For both demand and supply, there are three unknowns and two equations. The parameters of the structural equations are not recoverable. The identification problem posed by Brown and Rosen (1982) is not a result of the need to estimate marginal prices. Identification, in this case, remains a problem even when the price gradient is known. The identification problem is deeper, lying in the amount of nonlinearity in the hedonic and structural equations.

Of the three potential identification problems facing hedonic models, we have shown that the first two are merely surface reactions to the third. The simultaneity of demand and supply and the use of estimated prices in the structural equations are special problems in hedonics only to the extent that they reflect the problem of simultaneity between price and shift variables. The problem with single market data is that prices and exogenous structural shift variables vary together throughout the sample. In this section, we explore the assumptions about functional form which are necessary and sufficient to identify structural equations with data from a single market. By restricting the permitted functional form of the structural equations, the nonlinearity of marginal prices can be used to identify the price and shift parameters of both demand and supply. The identification approach must be used with great caution, however, because the true shape of supply and demand functions are often unknown and so the necessary restrictions may be unjustified.

Let us assume we observe a set of constant marginal prices $p_i(z)$ for a single good or characteristic z. The characteristic could be a typical measure of quality such as the number of bedrooms in a house or the horsepower of a car. As discussed by Rosen (1974), we assume the price gradient is the equilibrium of a multiplicity of supply and demand curves. Each actor is assumed to be a price taker (more precisely, a price gradient taker) in that the price gradient is determined exogenously to the actor. Consumers are assumed to maximize well-behaved utility functions subject to the budget constraint imposed by their income and all market prices (including the price gradient). Similarly, suppliers are assumed to maximize their profits subject to technology, input prices, and the price gradient (output price schedule). In addition to observing the price gradient, let us assume we observe the demand (y) and supply (w) shift variables of, respectively, each purchaser and producer interacting in this market.

As Hall (1973) has shown, maximization of utility subject to a nonlinear budget constraint is equivalent to maximization of utility with respect to a linear budget constraint which is tangent to the nonlinear constraint at the optimum bundle z^*. Assuming second order conditions are satisfied, the behavior of the consumer can be described in terms of a set of simultaneous equations:

$$p = F(z^*,y) \tag{3.7}$$

$$p = p_i(z^*). \tag{3.8}$$

The first equation is a traditional inverse demand function defined over a linear budget constraint. The second equation adjusts marginal prices to keep the individual upon the nonlinear budget constraint.[2] Together, these equations characterize a consumer's behavior subject to the price gradient $p_i(z)$. A parallel construction is clearly possible upon the supply side generating:

$$p = G(z^*,w) \tag{3.9}$$

$$p = p_i(z^*) \tag{3.10}$$

where G(z,w) is the inverse supply curve assuming constant output prices and (3.10) is the same price gradient as (3.8).

For the demanders and suppliers represented by (3.7) and (3.9) to have produced the price gradient (3.8) or (3.10), it must be true that

$$p_i(z^*) = F(z^*,y) = G(z^*,w). \tag{3.11}$$

For each observed level of z, the buyers should have the characteristics, y, which would generate a marginal willingness to pay of $p_i(z)$. Similarly, the sellers should be observed to have characteristics, w, for a marginal willingness to sell equal to $p_i(z)$. This consistency requirement (3.11) is the source of the identification problems inherent in a single market.

Given heterogeneous actors in the market and a single price gradient, the only consistent reason agents choose different bundles is because of their shift variable.[3] Let us assume that each shift variable y or w has a monotonic effect on demand or supply, respectively. Holding the price gradient constant and simply varying y (or w) would result in a monotonic relationship between z and the level of y (or w). For example, as income increases, consumers purchase more of each normal good and less of each inferior good throughout the range of observed incomes. Let us describe this expansion path in terms of a function $\phi(\cdot)$ and $\lambda(\cdot)$ for demand and supply respectively:

$$z = \phi(y) \tag{3.12a}$$

$$z = \lambda(w). \tag{3.12b}$$

Because $\phi(\cdot)$ and $\lambda(\cdot)$ are monotonic functions, their inverse must exist. Let us define this inverse as:

$$y = A(z) \tag{3.13a}$$

$$w = D(z). \tag{3.13b}$$

The solution to (3.11) is (3.13a) and (3.13b). The shape of A and D depend upon both the shape of the price gradient and the functional form of the underlying structural equations. Substituting (3.13a) and (3.13b) back into (3.11) provides a framework to analyze the identification issue:

$$p_i(z) = F(z,A(z)) = G(z,D(z)). \tag{3.14}$$

Intuitively, the problem with single market data is that exogenous shift variables and prices are functionally related. It is as though one chose a sample design so that for every increasing level of price there would be an increasing (or decreasing) level of the shift variables. Separating out the effect of prices from that of shift variables becomes difficult. For example, with one good and one shift variable, single market data can be represented by a single monotonic curve in three dimensional good, price, and shift variable space. An infinite number of structural equation surfaces could fit this single nonlinear curve. Further, even in the neighborhood of the observations, the set of consistent structural equations can have widely differing properties.

In order to analyze how nonlinearity can yield identification, it is necessary to characterize nonlinearity in concrete terms. Let us assume, therefore, that each function is a polynomial:

$$P_i(z) = \sum_{i=1}^{I} \gamma_i z^{i-1}$$

$$F(z,y) = \sum_{j,k}^{J,K} \beta_{j,k}\, z^{j-1}\, y^{k-1}$$

$$G(z,w) = \sum_{i,m}^{L,M} g_{i,m}\, z^{i-1}\, w^{m-1}$$

where γ_i, $\beta_{j,k}$, and $g_{i,m}$ are all constants and I, J, K, L, and M represent the number of nonzero terms in each expression. Let us further assume that A(z) and D(z) can be written:

$$y = A(z) = \sum_{n}^{N} a_n z^{n-1}$$

$$w = D(z) = \sum_{q}^{Q} d_q z^{q-1}.$$

Substituting the above expressions into (3.14) yields:

$$\sum_{i}^{I} \gamma_i z^{i-1} = \sum_{j,k}^{J,K} \beta_{j,k}\, z^{j-1} \left(\sum_{n}^{N} a_n\, z^{n-1}\right)^{k-1}$$

$$\sum_{i}^{I} \gamma_i z^{i-1} = \sum_{i,m}^{L,M} g_{i,m}\, z^{i-1} \left(\sum_{q}^{Q} d_q\, z^{q-1}\right)^{m-1} \quad . \qquad (3.15)$$

Since the above equations must hold for all levels of z, the coefficient for each term z^{i-1} on the left-hand side must equal the sum of the coefficients for the corresponding term of z on the right-hand side of (3.15a) and (3.15b). For example, associated with z^{I-1}:

$$\gamma_I = \beta_{J,K}\, a_N^{K-1} = g_{L,M}\, d_Q^{M-1}.$$

There is a separate demand and supply side equation for each power of z. Compressing this information in matrix notation yields:

$$\gamma = A\beta \text{ and} \tag{3.16a}$$

$$\gamma = DG \tag{3.16b}$$

where γ is a Ix1 vector, β is a JKx1 vector, A is a matrix IxJK, G is a vector LMx1 and D is a matrix IxLM.

The clue to the identification role of nonlinearity lies in (3.16a) and (3.16b). The parameters in γ are observable; they simply reflect the price gradient. The parameters in A and D are also known since these reflect the observable expansion path between y or w and z. What is unknown are the parameters in β and G. Solving (3.16a) and (3.16b) for β and G yields:

$$\beta = (A'A)^{-1} A'\gamma \tag{3.17a}$$

$$G = (D'D)^{-1} D'\gamma. \tag{3.17b}$$

A necessary condition for solving (3.17a) and (3.17b) is that there be as many equations as there are unknowns. Thus, for a unique solution, $1 > JK$ and $1 > LM$ for the demand and supply side, respectively. The number of nonzero terms in the price gradient must be equal to or greater than the number of nonzero terms in the structural equation.

A sufficient condition for solving β and G in (3.17a) and (3.17b) is that the number of linearly independent rows in A and D equal or exceed the number of parameters to be estimated. That is, the number of linearly independent parameters in the price gradient must exceed the number of parameters which must be estimated in the structural equations.

These simple results can easily be extended to incorporate vectors of characteristics or demand and supply shift variables. Correspondingly, nonlinearity can be measured by the increased number of parameters in the price and structural equations in these more complex models. For example, instead of the demand parameters β being JK, they could be expanded to $\sum_{i}^{N} JK_i$ with N characteristics.

Adding interaction terms among the characteristics would complicate the model further requiring even more parameters to be estimated. Interaction terms can add to the nonlinearity of either the price gradient or the structural equations. Each function could not only include single powers of each characteristic but also multiplicative terms amongst the characteristics. For example, the polynomial of each function could include all terms whose sum of exponents does not exceed a parameter, r. As an illustration, a price gradient with two characteristics and an exponent limit $r_p = 3$ would include the following terms:

$$z_1,\ z_1^2,\ z_1^3,\ z_2,\ z_2^2,\ z_2^3,\ z_1 z_2,\ z_1^2 z_2,\ z_2^2 z_1.$$

For any polynomial with N characteristics and r exponent limit, the number of terms would be: $rN + \sum_{i=1}^{N} (i-1) \sum_{j=1}^{r} (j-1)$.

The solution to this more difficult problem can be written in terms of equations (3.16) and (3.17) by redefining the individual vectors and matrices. A, γ, and β would have the following dimensions:

$$A: \quad r_\gamma N + \sum_{i=1}^{N} (i-1) \sum_{j=1}^{r_\gamma} (j-1) \; X r_\beta N + \sum_{i=1}^{N} (i-1) \sum_{j=1}^{r_\beta} (j-1)$$

$$\gamma: \quad 1 \times r_\gamma N + \sum_{i=1}^{N} (i-1) \sum_{j=1}^{r_\gamma} (j-1)$$

$$\beta: \quad 1 \times r_\beta N + \sum_{i=1}^{N} (i-1) \sum_{j=1}^{r_\beta} (j-1)$$

where r_γ is the exponent power of terms in the price gradient and r_β is the exponent power of terms in the demand equation. A parallel transformation would occur in the supply side. There would be a separate equation for each of the N characteristics in z.

The solution for β and G can be characterized by (3.17). The necessary condition is that the number of nonzero terms in the price gradient be equal to or greater than the number of nonzero terms in the structural equation. The sufficient condition is that the number of linearly independent nonzero terms in the price gradient exceed the number of terms needed for estimation in the structural equations.

To illustrate how nonlinearity can lead to identification, we reproduce the Brown and Rosen model but allow the marginal price gradient to be quadratic:

$$p_i(z) = \gamma_0 + \gamma_1 z + \gamma_2 z^2 \; .$$

As shown in Section 3.3, suppose the demand and supply curves are:

$$p_i(z) = \beta_0 + \beta_1 z + \beta_2 y \qquad \text{(demand)}$$

[illegible]

Utilizing (3.11) and the above equations, it is evident that A(z) and D(z) must be quadratic:

$$y = \frac{\gamma_0 - \beta_0}{\beta_2} + \left[\frac{\gamma_1 - \beta_1}{\beta_2}\right] z + \left[\frac{\gamma_2}{\beta_2}\right] z^2 \qquad (3.18)$$

$$w = \frac{\gamma_0 - G_0}{G_2} + \left[\frac{\gamma_1 - G_1}{G_2}\right] z + \left[\frac{\gamma_2}{G_2}\right] z^2 .$$

Given observations about y, w and z, this quadratic expansion path could be estimated:

$$y = q_0 + q_1 z + q_2 z^2 \tag{3.19}$$

$$w = d_0 + d_1 z + d_2 z^2.$$

For the data to be consistent with (3.18) and (3.19), it follows that:

$$\frac{\gamma_0 - \beta_0}{\beta_2} = q_0 \qquad \frac{\gamma_0 - G_0}{G_2} = d_0$$

$$\frac{\gamma_1 - \beta_1}{\beta_2} = q_1 \qquad \frac{\gamma_0 - G_1}{G_2} = d_1$$

$$\frac{\gamma_2}{\beta_2} = q_2 \qquad \frac{\gamma_2}{G_2} = d_2 .$$

With both demand and supply, there are three equations and three unknowns. The underlying shift parameters can be recovered from the observable data because the nonlinearity of the price gradient is at least as great as the number of nonzero parameters which had to be estimated in the structural equations.

Another common assumption made in the early hedonic literature is that all persons are the same. If all persons are the same, there are no structural parameters to estimate and the number of terms in the price gradient will always equal the number of terms in the structural equation. In fact, the structural equation will always be the price gradient in this case. Rare indeed are the circumstances where this is a legitimate assumption.

A more reasonable assumption was suggested by Quigley (1982). To sort between income shift effects and prices, Quigley suggests assuming an income or shift elasticity of one. With the resulting homothetic preference restriction, the information in a single market could be used to measure the residual price effect.

Contrary to Quigley's assertions, however, his approach does not generalize to more complicated preference maps. When income elasticities as well as price elasticities need to be estimated there are multiple solutions to the underlying demand or utility parameters. That is, identification is only assured when the income elasticity is chosen by assumption. It cannot be simultaneously estimated along with the price elasticity using single market data.

3.5 Conclusion

Section 3.4 illustrates the sufficient and necessary conditions for identification of structural equations using the nonlinearity of a single price gradient. If the number of terms in the structural equation is limited to the number of linearly independent terms in the price gradient, the parameters of the structural equation can be identified. Additional terms are possible from higher powers of each characteristic and also from interaction terms among the characteristics.

There is information about the behavior of consumers and suppliers in a single market. The information, however, is not as complete as in the multiple market case. Consequently, it is necessary to restrict the functional form of the structural equations to permit use of single market data. If such functional form restrictions can be justified (for example by being tested on mutliple market data), then single market data analysis could serve as a useful supplement to multiple market analysis. All too frequently, however, assumptions about functional form are made for convenience only. If the true functional form has too many parameters to be identified with data from a single market, arbitrary restrictions of functional form will produce arbitrary results. No matter how well the unidentified functional form fits the data, the results would not necessarily approximate the truth, even in the neighborhood of the observations. Although the choice of functional form for estimation purposes may or may not be a serious issue, the same choice of functional form to justify identification is always critical. Given how little is known about the true shape of structural equations and how important that information is to single market analyses, practitioners should be highly cautious about using single market data to reveal structural equations.

If possible, analysts should turn to multiple market examples, either intertemporal or cross-sectional. By varying the price gradients facing individuals, one can break the functional relationships $A(z)$ and $B(z)$ between prices and exogenous variables which plague single market data. In fact, it is only the existence of exogenous variation of price gradients which prevents a much larger set of papers in the labor, electricity, and urban literature from falling prey to the identification problem discussed in this paper.

There are several papers which have utilized multiple markets to properly estimate hedonic structural equations. Palmquist (1982) uses housing data from several cities to estimate the demand for housing characteristics. Mendelsohn (1980) uses workplace location to identify spatially separated housing markets for estimating the demand for housing characteristics. Brown and Mendelsohn (1984) use residential users to estimate the demand for recreation characteristics.

[illegible]

subdividing a single market into independent submarkets. For example, King (1976) and Strazheim (1973) attempt to estimate the demand for housing characteristics by assuming that different towns within a single metropolitan area are different markets. Unfortunately, the choice of whether to live downtown or in the suburbs is generally made precisely because of the housing characteristics. The assumption that these are independent markets will frequently be inappropriate. Single market identification cannot be corrected by arbitrarily subdividing the market into smaller submarkets.

CHAPTER 3

FOOTNOTES

1 School of Forestry and Environmental Studies and Department of Economics, Yale University. Many thanks go to K. E. McConnell for his administrative support and substantive comments. I would also like to thank Michael Hanemann for his helpful criticisms.

2 In addition to the marginal price effect, there is also an income effect associated with the change in inframarginal prices. For most examples, this income effect is small and for expositional simplicity it is omitted in the following discussion.

3 If all consumers are alike, the price gradient would reflect a compensated demand function. If all suppliers are alike, the price gradient would reflect an iso-profit supply function. If both consumers and suppliers are alike, only one bundle would be transacted. Although perhaps extreme, these assumptions provide an example of how demand and supply can be estimated by restricting the model.

CHAPTER 4

IDENTIFICATION OF THE PARAMETERS OF THE PREFERENCE FUNCTION: CONSUMER DEMANDS WITH NONLINEAR BUDGETS

K. E. McConnell and T. T. Phipps[1]

4.1 Introduction

The hedonic approach has become widely accepted as a method of modelling quality choice in a market where prices reflect quality. A problem which arises in practice with the hedonic technique is the recovery of information about preferences for the quality of goods. Solutions to this problem, the so-called identification problem, have evolved from the initial suggestion by Rosen that exogenous market supply will solve the identification problem to arguments by Diamond and Smith (1985) for the use of multiple markets.

Despite the evolution of solutions to the identification problem, there is still a good deal of uncertainty about the issue. This uncertainty exists in part because there is little agreement on criteria for identification, and perhaps more fundamental, there is seldom explicit discussion of precisely what is being identified. For example, Brown and Rosen (1982) tie the identification problem to definitional links between marginal prices and quality levels, but give little guidance as to the precise nature of the function being estimated. Quigley (1982) derives the structural equations from explicit utility maximization, but does not deal with the potential for underidentification in this context. Thus, while there are many contributions on the identification problem, they tend to be fairly diverse in their statement of the problem and their approach to solutions.

The purpose of this chapter is to explore the problem of identifying the parameters of hedonic models in a framework consistent with choice theory and the structure of preferences. Our point of departure is that empirical hedonic analysis using observations on individual purchases (prices and attributes of goods) is strictly a problem of consumer demand analysis with a [illegible] econometric model relating to consumers' choices. The advantage of deriving the econometric structure from the household's utility maximization problem is two-fold. First, by utilizing the household model, we see exactly what the endogenous variables are and where they come from. Second, by requiring the household's maximization system to fit into a traditional econometric model, we avail ourselves of the use of traditional econometric criteria for identification.

Deriving the econometric structure from the household's choice problem provides considerable unifying insight into the identification problem. Among the insights this approach allows are:

- the Rosen two step approach requires restriction assumptions about errors and preferences;
- parameters of the hedonic price equation as well as the preference function are subject to underidentification;
- successful estimation by maximum likelihood is evidence of identification;
- the linear hedonic price equation can be used in some cases in a single market setting.

An especial advantage of the approach of this chapter is that it allows us to assess identification of parameters in single and multiple markets with the same criteria.

The chapter proceeds along the following line. In the next section, we develop the structure of choice for households with nonlinear budget constraints. This section is crucial because there we show precisely what we are seeking when we solve the identification problem. In section 4.3, we explore identification in the single market, showing how various criteria for identification can be used. In section 4.4, we address the use of multiple market data. Section 4.5 concludes the chapter with pessimistic arguments about the prospect of recovering parameters of preferences in hedonic markets.

4.2 The Structure of the Problem

In this section, we attempt to give a clear statement of the identification problem in hedonic markets and show briefly how others have addressed and solved the problem. The analysis assumes the existence of a hedonic market where buyers and sellers compete for the purchase or sale of a good with several attributes. Assumed measurable, these attributes are denoted z. The existence of this market implies a hedonic price equation:

$$p = h(z;\gamma) \qquad (4.1)$$

where p is the price of a unit of the good, z is a K-dimensional vector of attributes of the good, and γ is a vector of parameters which describe the function h. This equation gives the amount households expect to pay and firms expect to receive for units of the good characterized by the attribute vector z. Perfect competition is assumed, i.e., buyers and sellers treat the hedonic price <u>function</u> as given. They cannot influence the parameters γ but they can influence the price by the selection of z.

Our interest is in preferences for attributes. It is assumed households have a well defined preference function, designated $U(x,z;\beta)$ where x is a Hicksian bundle with a unit price and β is a vector of parameters describing preferences. Households choose levels of the vector z and the composite commodity x to maximize $U(x,z;\beta)$ subject to the budget constraint $y = x + h(z;\gamma)$, where y is household income. Equilibrium conditions for the optimal choice of the attribute vector by the household include

$$\frac{\partial h(z;\gamma)}{\partial z_i} = \frac{\partial U(x,z;\beta)/\partial z_i}{\partial U(x,z;\beta)/\partial x} \qquad i = 1,K. \qquad (4.2)$$

This condition states that the marginal price of the i^{th} attribute equals the marginal rate of substitution between the i^{th} attribute and the numeraire good. Much of the hedonic literature presumes that solutions to the equilibrium condition (4.2) exist in the form of direct or inverse functions for z and focuses on the estimation of the presumed demand functions.

We now have sufficient structure to give a clear statement of the identification problem. Given observations on household purchases,

> Can we recover the vector of parameters, β, which describes household preferences?

The identification problem is solved when we have enough of the parameters of $U(x,z;\beta)$ to calculate the change in a household's welfare from an exogenous change in the attribute bundle, given income.

The identification problem in hedonic markets is different from the problem typically encountered in simultaneously estimating supply and demand functions. The general problem of estimating supply and demand functions arises with three types of data sets. First, one can use aggregate market data to estimate these functions. An identification problem arises because market price is simultaneously determined with aggregate quantity. Second, one can estimate parameters of supply and demand using individual data on quantities and prices for firms or households when individual actors are price takers. There is no identification problem in this setting because price is exogenous to the individual quantities chosen. Third, one can estimate behavioral functions from disaggregate data when individuals are not price takers. In this case, where there are monopolistic or monopsonistic elements, the same type of identification problem found in hedonic models is encountered.

The consequence of the identification problem is that parameters of preference (β's) are confused with parameters of the hedonic price equation (γ's). This is similar to the problem in separating tastes and technology in the household production function.[2] For econometric purposes, the structures of hedonic models and household production models are quite similar. The most important difference between the two structures is that the budget constraint from the household production function must be convex, because it results from a household minimization problem. The hedonic price equation is not constrained to be convex by any market forces, and as we discuss in [illegible] when it comes to measuring welfare changes.

We have defined the identification in hedonic markets to be the recovery of the parameters of the preference function. Since the literature typically discusses identification of the parameters of demand functions, we explore briefly the distinction.

Recall the equilibrium conditions for an optimum as:

$$\partial h(z;\gamma)/\partial z_i = \frac{\partial U(x,z;\beta)/\partial z_i}{\partial U(x,z;\beta)/\partial x} \qquad i = 1,K \qquad (4.3)$$

For ease of notation, denote

$$m_i(x,z;\beta) \equiv \frac{\partial U(x,z;\beta)/\partial z_i}{\partial U(x,z;\beta)/\partial x}$$

The equilibrium condition requires marginal price to equal the marginal rate of substitution between the attribute and the numeraire good. We may denote $m_i(x,z;\beta)$ as the marginal rate of substitution function. It is this function which is the so-called demand function for attributes or the "hedonic demand function".

The marginal rate of substitution function differs from the inverse demand function. Further, when the budget constraint is nonlinear, neither the direct nor inverse Marshallian demand functions exist as solutions to the consumer's maximization problem. These points are crucial because they bear on estimation, interpretation and welfare measurement. First, consider the difference between the m_i function for the hedonic problem and the i^{th} inverse demand function from a traditional linear budget constraint problem. In the traditional problem, the consumer chooses levels of a K-dimensional vector x at constant prices p in order to

$$\max_x \{U(x) \mid px - y = 0\}.$$

Then the inverse demand functions are (where $U_i \equiv \partial U/\partial x_i$)

$$\frac{p_i}{y} = U_i/\sum_j U_j x_j \qquad i = 1,K \qquad (4.4)$$

by Wold's theorem. This problem has prices as parameters and has been completely solved. In contrast, the marginal rate of substitution conditions for the same consumer are

$$p_i/p_j = U_i/U_j \qquad i = 1,K.$$

These are equilibrium conditions which have not been solved to eliminate the [illegible] inverse demand functions have entirely different implications for estimation and welfare calculation.

For the individual household, the nonlinear hedonic price function creates a nonlinear budget constraint. There are two consequences of a nonlinear budget constraint for the utility-maximizing or cost-minimizing household. First, Marshallian and Hicksian demand curves as traditionally conceived, where price taking consumers choose quantities (utility or income held constant) do not exist. These demand concepts depend entirely on the linear budget constraint or prices-as-parameters paradigm. Second, the solution of the first order or equilibrium conditions gives quantities demanded of the attributes as a function of the parameters of the hedonic price equation as well as income and other exogenous variables.

Traditional demand functions, both Marshallian and Hicksian, rely on the happy coincidence that some of the parameters of the consumer's maximization problem are the prices of the goods. It is always correct to solve for optimal quantities of the goods (or attributes) as functions of parameters. But only when these parameters are also per unit prices will the traditional demand functions, with all their well known properties, emerge. The failure of these traditional concepts to hold when the budget constraint is nonlinear can be shown intuitively in two ways. First, one can attempt the mental experiment of asking: If the price were \$p, how much would be consumed? It is clear that asking this question when the budget constraint is nonlinear requires one to know already how much is being chosen. The absence of a traditional Marshallian or Hicksian demand when the budget constraint is nonlinear is analogous to the absence of a supply curve for a monopolist. Both concepts require that prices be parameters. When prices are not parameters, neither function exists. Second, one can construct examples, (as in the appendix to this chapter), given preference and cost functions, which show that quantities depend on parameters and not on average or marginal prices. Similarly, one can also show that well-behaved inverse demand functions are not defined for nonlinear budget constraints.[3] (These results on Hicksian and Marshallian demand functions are developed in more detail in Bockstael and McConnell, 1983.)

Consider the equilibrium conditions (4.3) again. In principle, when combined with the budget constraint, these conditions can be solved for z and x. If the derivative on the left hand side of (4.3) were constant, i.e., the hedonic price equation is linear, then the solution of (4.3) would be a traditional demand function. The existence of a traditional demand function, with prices as parameters, is assumed in most hedonic work which pursues Rosen's second step. However, when $h(z;\gamma)$ is nonlinear, the parameters and exogenous variables on which z depends are income and the parameters of the hedonic price equation, so that the solutions for quantities chosen are:

$$z = D(y,\gamma;\beta) \quad (4.5)$$

$$x = D_x(y,\gamma;\beta) \quad (4.6)$$

[illegible] D and D [illegible] They are Marshallian only in that they depend on parameters. But they do not depend on prices.

The solution for z is a demand function in that it describes how choices depend upon parameters. Because prices are not parameters, they are not arguments in (4.5) and (4.6). When the hedonic price equation changes, the vector γ changes, and households respond. Expressions (4.5) and (4.6) are reduced form equations. Estimating these equations allows one to make predictions of z and x. But successful estimation of (4.5) and (4.6) solves the identification problem only when the parameters of the hedonic equation and the preference function can be deduced from the reduced form parameters.

The conclusion of this section concerns the question "what are we seeking?" when we attempt to identify demand structure in hedonic models. The answer is that we are seeking to identify the marginal rate of substitution functions. These functions are demand relations only in the sense that they equal marginal price at optimum. The true demand functions can be solved for only rarely; hence, the equilibrium conditions must be estimated. There are at least two practical consequences of this result. For estimation purposes, the structural equation must integrate back to a quasi-concave function, ruling out most polynomials. Further it will in general include the hedonic price as an argument. Second, when computing welfare changes, one must either start with a utility function and derive the implied marginal rate of substitution functions or start with the marginal rate of substitution functions and derive the appropriate welfare functions. We have also explained why we have framed the problem as one of recovering the parameters of the preference function. These parameters are embodied in the marginal rate of substitution functions. Traditional Marshallian and Hicksian direct and inverse demand functions do not exist as solutions when the hedonic equation is nonlinear.

In the following sections we discuss the identification problem for two kinds of hedonic models. The first deals with simultaneous estimation of the hedonic price equation and the marginal rate of substitution functions for a single market. This arises when both hedonic parameters and preference parameters are estimated from the same set of transactions data. Identification criteria are derived for models that are linear and nonlinear in parameters. Within the single market, we consider two special cases: the case of the linear hedonic equation and the case when the hedonic parameters are available from an alternative source, and only the preference parameters are estimated. The second kind of hedonic model concerns identification from multiple markets.

The criteria we develop are based on the econometric theory of identification of the parameters of linear and nonlinear systems. Hence, the identifiability of a system will be determined by the restrictions we impose, i.e., homogeneous and nonhomogeneous parameter restrictions, across-equation parameter restrictions, and the specification of the functional form of the hedonic and marginal rate of substitution equations. It is shown that, unlike the traditional problem of identifying supply and demand equations by [illegible]
identify hedonic models. Identification of the parameters of hedonic models generally involves the imposition of untestable restrictions on the functional forms of the equations of the system.

4.3 Single Market Approaches to Identification

This section presents an approach to hedonic models which views the marginal rate of substitution functions as part of a system. Here we are interested in hedonic analysis of observations on prices and attributes of goods which come from sales transactions. We can then be confident that when the traditional hedonic story is told, prices and attributes will be jointly dependent. This section presents an approach to identification that brings us closer to the question of whether it is possible to identify the parameters of concern.

Let us construct the econometric system. The maximization problem, when z is a scalar,

$$\max_{x,z} \{U(x,z;\beta) \mid y-h(z;\gamma) - x = 0\}$$

has three first order conditions

$$U_X(x,z;\beta) = \lambda$$

$$U_Z(x,z;\beta) = \lambda \partial h/\partial z$$

$$y - h(z;\gamma) - x = 0$$

and three choice variables (x,z,λ). The ratio of the first to the second yields

$$\partial h/\partial z = m(x,z;\beta)$$

where $m(\cdot)$ is the marginal rate of substitution function. Defining $p \equiv y - x$ and substituting $x = y - p$ for x wherever it appears yields two unknowns (z,p) in two equations

$$p = h(z;\gamma)$$

$$\partial h/\partial z = m(y - p,z;\beta).$$

These are the two structural equations of the consumer's optimum. For the K attributes case, there would be K+1 equations and K+1 unknowns, but the basic arguments remain. In analyzing transactions data involving prices and attributes of goods, we should treat these variables as jointly dependent.

This characterization of the structure is significantly different from the standard hedonic literature. Typically, both a supply function and a demand or marginal rate of substitution function are specified, with z and $\partial h/\partial z$ as endogenous. (See for example, S. Rosen, 1974, or Brown and H. Rosen, 1982.) At the margin, the hedonic model is analogous to the typical market model of supply and demand. While it is intuitively appealing to utilize the market model of supply price and demand price, it is misleading in the household case with nonlinear budget constraints. In the context of the individual consumer's choice, consumers are price schedule takers and we may safely ignore the modelling of sellers' decisions. In this context, knowledge of z and $\partial h/\partial z$ does not allow the computation of utility without further information. In fact, knowledge of $\partial h/\partial z$ is of no particular value to the consumer. It cannot be used to predict consumer choices or utility.

Accepting that z and either p or x are endogenous, one naturally asks: why not solve for z and x in terms of the exogenous variables? The answer is that it is easier said than done. Solving for z and x requires severe restrictions on the preference functions and the hedonic price equation. Moreover, it is often not possible to solve for z and x given even the simplest preference function and nonlinear hedonic price equations. (See the example in the appendix to this chapter.)

Because we generally cannot solve for the endogenous variables, we are forced to estimate the equilibrium conditions. When we are analyzing transactions data with observations on purchase price and attributes, we can capture the econometric spirit of the choices facing the consumer by specifying the following system:

$$p = h(z;\gamma) + \varepsilon_1 \qquad (4.7)$$

$$\partial h(z;\gamma)/\partial z_i = m_i(y - p, z;\beta) + \varepsilon_{i2} \qquad i = 1,K \qquad (4.8)$$

where γ is the unknown vector of parameters of the hedonic price function and β is the unknown vector of parameters of the preference function. The endogenous variables are p and the vector z, and the functions m_i are the marginal rate of substitution functions. Because the hedonic model is not customarily written as in (4.7) and (4.8), there is little discussion of the probability densities of the ε's. Specifying the error structure in hedonic models should be an integral part of model construction. The error term in the hedonic equation may arise from errors in measurement, unobserved or omitted variables, and approximation errors due to lack of knowledge of the true functional form of the hedonic equation. The error term in the marginal rate of substitution equations may arise from the same type of misspecification encountered with the hedonic equation, though we have the additional problem of unobserved variation in tastes across households. The errors are econometrician's errors rather than stochastic elements in household behavior. No prior restrictions are obvious, so it makes sense to specify them as having mean zero and constant variance.

The general hedonic model to be estimated is the system (4.7) and (4.8). For identification, it is necessary to determine whether:

- The parameters of the hedonic price equations are identifiable;
- The parameters of the marginal rate of substitution function are identifiable;
- The parameters of the whole system are identifiable.

The focus of the identification debate has been whether parameters relating to individual behavior (here the marginal rate of substitution function) can be identified. It perhaps makes more sense to ask whether the hedonic structure or the system as a whole can be identified. Several dimensions of the hedonic model warrant attention.

- The model almost certainly will be nonlinear in variables;
- For most preference functions, the model will also be nonlinear in parameters;
- There may be shared parameters in different equations or cross-equation parameter constraints.

4.3A. Models Linear in Parameters But Not in Variables

A source of difficulty in identifying hedonic models is nonlinearity. For the case of models which are linear in parameters, however, identification criteria are well established. Suppose there are M equations and endogenous variables. When there are K attributes, then the system (4.7) and (4.8) has M = K + 1. Assuming there are no implied equations, we let the system be written

$$Aq(w) = \varepsilon$$

where A is the M by N parameter matrix and q(w) is the N element vector of basic endogenous variables, exogenous variables, and functions of endogenous variables, which are labelled additional endogenous variables. Let Φ_i be the matrix of prior homogeneous restrictions for the i^{th} equation. With no implied equation in the system, the necessary condition for identifiability of the i^{th} equation is

$$\text{rank } (\Phi_i) \geq M - 1 \qquad (4.9)$$

when a parameter has been normalized. The necessary and sufficient condition is

$$\text{rank } (A\Phi_i) = M - 1. \qquad (4.10)$$

The caveat that the conditions hold for equations with a normalized parameter is critical, for the marginal rate of substitution equation will be unnormalized of necessity. Normalization of a parameter in the marginal rate of substitution function in effect determines the relative value of coefficients in the utility function, and in many cases places quite restrictive assumptions on tastes. For example, for one attribute, when preferences are given by $U(x,z) = \beta_1 \ln z + \beta_2 \ln x$, the marginal rate of substitution function is $(\beta_2/\beta_1)\, z/x$. A normalization of $\beta_2/\beta_1 = 1$ determines <u>all</u> of tastes. No estimation is then necessary.

When there are no normalized parameters, the necessary condition for the identification of the i^{th} equation is

$$\text{rank } (\Phi_i) \geq M \qquad (4.11)$$

[illegible]

$$\text{rank } (A\Phi_i) = M. \qquad (4.12)$$

Criteria (4.9) and (4.10) can be used for the hedonic price equation, while criteria (4.11) and (4.12) are suitable for the marginal rate of substitution equation.[4] Observe that by characterizing the hedonic price function and the marginal rate of substitution functions as the structure with p and z jointly endogenous, we uncover the possibility that the hedonic price equation as well as the marginal rate of substitution equation will be underidentified. This topic is explored in the following chapter.

It is revealing to utilize these criteria in the one attribute example discussed by Brown and Rosen (1982). This example is inconsistent with the spirit of Section 4.3 in that it does not integrate back to a quasi-concave utility function, nor does it contain p as an argument. However, it is useful because of its widespread consideration in the literature. Let

$$h(z;\gamma) = \gamma_0 + \gamma_1 z + \gamma_2 z^2/2 + \varepsilon_1 \tag{4.13}$$

and

$$m(y,z;\beta) = \beta_0 + \beta_1 z + \beta_2 y + \varepsilon_2 \ . \tag{4.14}$$

Then $q(w) = [p\ z\ 1\ y\ z^2/2]^T$ and

$$A = \begin{bmatrix} 1 & -\gamma_1 & -\gamma_0 & 0 & -\gamma_2 \\ 0 & \gamma_2-\beta_1 & \beta_0-\gamma_1 & -\beta_2 & 0 \end{bmatrix}$$

with

$$\phi_1 = [\ 0 \quad 0 \quad 0 \quad 1 \quad 0\ ]^T$$

and

$$\phi_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$$

where the T indicates transposition. Both conditions (4.9) and (4.10) are satisfied for the hedonic price equation:

$$\text{rank } (\phi_1) = \text{rank } (A\phi_1) = 1.$$

When we apply criterion (4.11) to the unnormalized rate of substitution equation (4.14), we see that the rank $(\phi_2) = 2$, so that the necessary condition holds. However, applying the necessary and sufficient conditions yields

$$\text{rank } (A\phi_2) = \text{ rank} \begin{bmatrix} 1 & -\gamma_2 \\ 0 & 0 \end{bmatrix} = 1 < M = 2$$

so that in fact the marginal rate of substitution equation is not identified. This application of the formal criteria for identification leads to the same results as Brown and H. Rosen's analysis of S. Rosen's two-step approach.

The standard linear restriction criteria developed by Fisher and extended to systems nonlinear in variables work as long as the constraints are simply written. However, when more complicated information becomes available, these criteria are not applicable. Such information becomes available when the hedonic price function is known to be more complicated. For cases, which are still linear in parameters, the work by Wegge (1965) provides the basis for identification. Wegge's criteria are similar in spirit to those of Fisher, but allow for cross-equation parameter constraints and nonlinear constraints.

Consider a one attribute example. Let the hedonic price function be

$$p = \gamma_0 + \gamma_1 z + \gamma_2 z^2/2 + \gamma_3 z^3/3 + \varepsilon_1. \qquad (4.15)$$

With the same marginal rate of substitution function as in (4.14), the equilibrium condition is

$$\gamma_1 + \gamma_2 z + \gamma_3 z^2 = \beta_0 + \beta_1 z_1 + \beta_2 y + \varepsilon_2. \qquad (4.16)$$

Utilizing the sufficiency criterion in (4.10) for (4.15) we find that rank $(A\phi_1) = 1$ so it is identified. Applying criterion (4.12) to the unnormalized (4.16) we see that rank $(A\phi_2) = 1$, so that it is not identified. We have added information which should help us distinguish between the two structural equations, but the standard criteria imply that the second equation is still not identified.

The intuitive explanation of this result comes from observing that in the system (4.15) and (4.16) there is an exact across-equation constraint. If we write the A matrix

$$A = \begin{bmatrix} 1 & -\gamma_1 & -\gamma_0 & -\gamma_2 & 0 & -\gamma_3 \\ 0 & \gamma_2-\beta_1 & \gamma_1-\beta_0 & 2\gamma_3 & -\beta_2 & 0 \end{bmatrix}, \qquad (4.17)$$

we see that $a_{24} + 2a_{16} = 0$. The identifiability of this system can be determined by Wegge's criterion. Strictly speaking, Wegge's results apply to systems linear in variables. In most cases we can harmlessly convert nonlinear systems to linear systems by substituting polynomial functions of exogenous variables for the additional endogenous variables. Our concern is to determine whether the two equations are observationally equivalent. Let T be any nonsingular M by M matrix and let vec (TA) be the vector created by taking TA one row at a time. If the two equations are observationally equivalent, constraints on A will also hold on TA. Let

$$\phi_i(\mathrm{vec}(A)) = 0 \qquad i = 1,R$$

be the vector of constraints, including normalizations, across equation parameter constraints, and homogeneous restrictions, where R is the number of such constraints. Define the matrix J as

$$J(T) = \frac{\partial\phi_i(\mathrm{vec}(TA))}{\partial \mathrm{vec}(T)} \qquad i = 1,R. \qquad (4.18)$$

Then a sufficient condition for the identification of the system is that

$$\mathrm{rank}\ (J(I)) = M^2 \qquad (4.19)$$

where I is the M by M identity matrix (see Wegge, Theorem II, p. 71). For Wegge's results, the constraints need not be linear or homogeneous.

The constraints that are implicit in the A matrix in (4.17) are

$$
\begin{aligned}
\phi_1 &: a_{11} - 1 &= 0 \\
\phi_2 &: a_{15} &= 0 \\
\phi_3 &: a_{21} &= 0 \\
\phi_4 &: 2a_{16} + a_{24} &= 0 \\
\phi_5 &: a_{26} &= 0.
\end{aligned}
\qquad (4.20)
$$

Computing J(I) gives

$$
J(I) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\gamma_2 & 2\gamma_3 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -2\beta_3 & -\gamma_2 & 2\gamma_3 \\ 0 & 0 & -\gamma_3 & 0 \end{bmatrix}. \qquad (4.21)
$$

Denote by J^* the matrix derived by deleting the first row of J. Then we find that

$$\det J^* = -4\gamma_3^3$$

implying that the rank of J is $4 = M^2$. Hence the sufficient condition holds for this system to be identified. Note that the requirement that $\gamma_3 \neq 0$ is quite intuitive because when $\gamma_3 = 0$, we have the model given by (4.13) and (4.14), which we have already shown to be underidentified. An extension of this system to several attributes, while maintaining the basic functional forms, will show that the hedonic price equation will no longer be identified, a result discussed in Chapter 5.

The conditions can be usefully applied in practice and can be easily generalized to the setting where there are several endogenous variables. The restrictions needed for identifying the marginal rate of substitution equations depend on the hedonic price function. In practice, the hedonic equation will be specified empirically, typically using Box-Cox techniques. This leads to nonlinearity.

4.3B <u>Models Nonlinear in Parameters</u>

While nonlinear analytic functions may be approximated as closely as desired by polynomials linear in parameters, many models are inherently nonlinear in the parameters. Further, specifying the functions as polynomials obscures the basic concavity or convexity which economic functions typically possess. Polynomials cannot in general be integrated back to quasi-concave preference functions. For example, the marginal rate of substitution function for the preference function

$$U = \beta_1 \ln(z - \beta_2) + \beta_3 \ln x$$

is

$$m(y-p, z; \beta) = \beta_1 y/\beta_3(z - \beta_2) - \beta_1 p/\beta_3(z - \beta_2)$$

which is nonlinear in β_2.

Hence, it is important to examine the conditions for identifying this class of model. The approach used is that of Rothenberg (1971) and Bowden (1973). In addition to providing necessary and sufficient conditions for identification of a wide class of parametric models, their approach links the existence of maximum likelihood parameter estimates with identification, which may have some practical applications.

The identification conditions have been stated most generally by Bowden. Let • be the vector of parameters to be estimated. (In the hedonic context • = (γ,β).) A sufficient condition for local identification is that the information matrix have full rank when evaluated at the true parameter point (•*). (See Bowden, section 3.) The necessary condition requires that, when •* is a locally identified regular point,[5] the information matrix possess full rank at •*. The nonsingularity of the information matrix is more useful in practice than in testing for identification on an a priori basis.

The nonlinearity in parameters makes the criterion difficult to apply analytically. When the model is nonlinear in parameters, it would be most unlikely for the first order conditions to be linear in •*. Hence solving for •* typically requires numerical methods. Without explicit solutions for •*, it is not generally possible to determine analytically the rank of the information matrix.

The requirement that a locally identified system possess a nonsingular information matrix has limited usefulness. From the perspective of maximum likelihood methods, the ability to obtain unique parameter estimates is sufficient to demonstrate local identifiability. When a well formulated model has been estimated using maximum likelihood methods, one can argue that the identification problem has been solved. However, the dimensionality and [illegible]

4.3C The Linear Hedonic Price Equation: A Special Case

Research on hedonic models uniformly dismisses the case of a linear hedonic price equation in a single market. There are good conceptual arguments against linearity. It implies that repackaging is possible. There is good reason to believe that two six-foot Cadillacs don't make a twelve-foot Cadillac. Intuitively, it means that an individual can buy unlimited quantities of a single attribute without raising its marginal price. Practically, a linear hedonic price function implies no variation in marginal prices. When the marginal price is endogenous, there is no variation in one of the endogenous variables. However, when we recognize that p and z are jointly endogenous, the linear hedonic price equation is no longer a hopeless case.

In the following, we show that it is possible to recover preference parameters from a single market's data, even when h(z) is linear. The purpose of this example is not to provide new and practical approaches. Rather it is presented as an illustration of potential gains from characterizing z and p as endogenous.

Consider the system (which is again inconsistent with what utility maximization tells us about the marginal rate of substitution function but is a useful example)

$$p = \gamma_0 + \gamma_1 z + \varepsilon_1 \tag{4.22}$$

$$\gamma_1 = \beta_1 z + \beta_2 y + \varepsilon_2. \tag{4.23}$$

The parameter matrix is

$$A = \begin{bmatrix} 1 & \gamma_1 & -\gamma_0 & 0 \\ 0 & -\beta_1 & \gamma_1 & -\beta_2 \end{bmatrix}.$$

In this model, there is an across equations parameter constraint ($a_{12} - a_{23} = 0$). Hence we can use Wegge's Theorem II (equation 4.19 above). There are four constraints

$$\begin{aligned} \phi_1 &: a_{11} - 1 &= 0 \\ \phi_2 &: a_{14} &= 0 \\ \phi_3 &: a_{21} &= 0 \\ \phi_4 &: a_{12} + a_{23} &= 0. \end{aligned}$$

Computing J(I) as given in (4.19) yields

$$J(I) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_1 & -\beta_1 & -\gamma_0 & \gamma_1 \end{bmatrix}$$

which has rank = 4 = M^2 because $\det(J(I)) = \beta_2\gamma_1 \neq 0$. Hence, we can obtain some information about preferences even when the hedonic price equation is linear. This information exists because consumers with different incomes and equal prices purchase different levels of attributes. Information about preferences can come from observing income effects as well as price effects.

4.3D Model Where the Parameters of the Hedonic Price Are Known

For a variety of practical reasons, the researcher may wish to estimate the hedonic price equation and the marginal rate of substitution functions separately. As an example, one may have a hedonic price equation estimated from the Annual Housing Survey. This source of information would be different from the individual transaction data and would suggest a different econometric structure. A reasonable structure would be

$$\partial h(z;\gamma)/\partial z_i = m_i(y,z;\beta) + \mu_i \qquad i = 1,K \qquad (4.24)$$

where now the endogenous variables are z_i, $i = 1,K$ and p is taken as exogenous. In expression (4.24) the γ's are known numbers.

The one attribute case is illustrative. Consider the quadratic hedonic price function - linear marginal rate of substitution function which gives the equilibrium condition

$$\gamma_1 + \gamma_2 z = \beta_0 + \beta_1 z + \beta_2 y + \varepsilon \ .$$

Solving for z gives the reduced form equation:

$$z = \pi_0 + \pi_1 y + \mu \ ,$$

where

$$\pi_0 = \frac{\beta_0 - \gamma_1}{\gamma_2 - \beta_1}$$

$$\pi_1 = \frac{\beta_2}{\gamma_2 - \beta_1} \ .$$

We have three coefficients (β_0, β_1, β_2) to recover, but only two reduced form parameters from which to find them. Hence we cannot identify the β's as m is specified. Prior information can obviously be useful in identifying the β's, even in the single equation case. For example, suppose that the marginal rate of substitution function is given by

$$m(y,z) = \beta_1 z + \beta_2 y + \varepsilon_2;$$

i.e. $\beta_0 = 0$. Then the reduced form remains the same but the β_i may be recovered from the relationships

$$\beta_1 = (\gamma_0 + \gamma_2\pi_0)/\pi_0$$

$$\beta_2 = \pi_1(\gamma_2 - \beta_1) \ .$$

Of course this method of identification requires belief in the maintained hypothesis that β_0 is zero, which is not testable nor does it have any obvious behavioral implications. It is thus a good example of the kinds of restrictions needed in solving the identification problem. We can generally make assumptions analogous to $\beta_0 = 0$, but we will rarely have good economic reasons for such assumptions. However, the approach is easy to use for one attribute. As long as we can solve for z, estimate the reduced form parameters, and recover estimates of β from the reduced form parameters π and γ, then we can identify β.

The heart of the matter is of course the multi-attribute case, when the γ's are known constants. The system is

$$\begin{aligned} h_1(z;\gamma) &= m_1(y,z;\beta) + \mu_1 \\ &\ \ \vdots \\ h_K(z;\gamma) &= m_K(y,z;\beta) + \mu_K \,. \end{aligned} \tag{4.25}$$

If both the hedonic price equation and the utility function are strongly separable in the attributes and the errors uncorrelated ($E\mu_i\mu_j = 0$, $i \neq j$), each equation in the system (4.25) can be treated separately. This would be analogous to the one attribute case, but highly unlikely.

In general, we must treat (4.25) as a system of K equations in K endogenous variables. As in the previous analysis, it is useful to think of two cases. First, when (4.25) is linear in the β's, some form of least squares may be applied. Second, when (4.25) is nonlinear in β's, ML methods are required. In either case, what is the role of the γ's?

Consider first the linear-in-parameters case. In that case, the h_i are nonlinear functions of endogenous variables, and may be considered additional endogenous variables. As long as the h_i are not linearly dependent, there are K-1 exclusion restrictions for each equation. Further assuming the coefficient on h_i is known (and equals unity) only K-1 restrictions are required for identification. Consider a case where $K = 2$ and and m_i are linear in parameters and endogenous variables:

$$h_1(z;\gamma) = \beta_{01} + \beta_{11}z_1 + \beta_{12}z_2 + \beta_{13}y + \mu_1 \tag{4.26a}$$

$$h_2(z;\gamma) = \beta_{02} + \beta_{21}z_1 + \beta_{22}z_2 + \beta_{23}y + \mu_2 \tag{4.26b}$$

Given z, h_1 can be computed because γ is known. Hence its coefficient is unity. Without changing the substance of the problem, we can divide each equation in (4.26) by β_{ii}. This yields the coefficient matrix

$$A = \begin{bmatrix} 1 & \beta^*_{12} & \beta^*_{01} & -1/\beta_{11} & 0 & \beta^*_{13} \\ \beta^*_{21} & 1 & \beta^*_{02} & 0 & -1/\beta_{22} & \beta^*_{23} \end{bmatrix}$$

(columns: z_1, z_2, 1, h_1, h_2, y)

where $\beta^*_{ij} = \beta_{ij}/\beta_{ii}$. The restriction matrix for the first equation is

$$\phi_1 = [0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0]$$

implying that $\text{rank}(\phi_1) = \text{rank}(A\phi_1) = 1$, so that both necessary and sufficient conditions for identifying equation (4.26a) and (4.26b) are met.

The successful application of nonlinear 2SLS in practice depends upon how linearly independent h_1, h_2 and z are. Thus, for example, if the hedonic price equations were quadratic, as in the Brown and Rosen case, the marginal prices would be linear, h_i would be perfectly correlated with the right hand side of (4.26) and 2SLS not feasible. If they are quite collinear, then while identification holds formally, actual parameter estimates will be imprecise. Further, nonlinearity in $h(z;\gamma)$ is not sufficient to guarantee that h_i and h_j are linearly independent. For example, suppose that we have a Box-Cox transformation of a linear function of z's:

$$h(z) = [\bullet\gamma z - 1]^{1/\bullet}.$$

For this case $\gamma_i p^{\bullet - 1} = h_i$ so that

$$h_j = \frac{\gamma_j}{\gamma_i} h_i .$$

That is, the h_i are not linearly independent of each other, regardless of the value of $\bullet$, which determines the nonlinearity of h. No restrictions would be provided by this functional form.

The h_i also play a role in identifying the β in non-linear systems, though the role is less straightforward because ML methods are needed. To get some insight into ML models, suppose that the μ_i are distributed as independent normals with mean zero and variance σ_i^2. Since our observations concern the vector z, we must transform from μ to z. The log-likelihood for the t^{th} observation would be

$$c + \ln J(t) - \sum_{i=1}^{K} \sigma_i^2 \{h_i(z;\gamma) - m_i(y,z;\beta_i)\}^2/2$$

where

$$J(t) = \det \begin{bmatrix} h_{11} - m_{11} & h_{1K} - m_{1K} \\ h_{1K} - m_{1K} & h_{KK} - m_{KK} \end{bmatrix}.$$

Note that $h_{ij} - m_{ij} = \frac{\partial}{\partial z_j}(h_i(z;\gamma) - m_i(z,y;\beta))$ will depend upon β as long as the m_i are not separable in the z. Hence the derivatives of the likelihood function with respect to β will depend on the h_{ij} functions. The precise way in which h_i influences the log-likelihood can only be determined on a case-by-case basis. But the essence of the matter is the choice of endogenous variables. It can be shown that by designating z_i as endogenous and $h_i(z;\gamma)$ as nonbasic endogenous, the Jacobian of transformation has the effect of moving the estimates of β's away from those that minimize the squared error $(h_i-m_i)^2$. Thus, while the practical effects of the ML criterion, nonsingularity of the information matrix, are not great, framing the problem as ML demonstrates the role of the h_i. The choice of endogenous variable influences the parameter estimates. The endogenous variables which accord most with consumer choice are the z_i.

The situation where the γ's are estimated with error is the case considered by Epple (1982) and by Bartik (1983). In that case, we consider the realistic situation where the hedonic price equation is misspecified by omitting attributes of the good. By the solutions (4.5) and (4.6) we know that any attribute is a function of income, and hence correlated with income. Thus, for example, omitting the attribute view from a sufficiently nonlinear hedonic price equation will cause error in the marginal price to be correlated with the view, and hence with income. (In this case, income can stand for a whole vector of socioeconomic characteristics without changing the argument.) Thus misspecification of the hedonic price equation will make errors (μ_i) correlated with income (y) and seriously undermine any attempt to recover the β's.

4.4 Multiple Markets

Several researchers (Diamond and Smith (1983), Parsons (1985), Palmquist (1984)) have concluded that the use of multiple market data holds the most promise for recovering preference parameters for hedonic models. Multiple markets might exist in housing, for example, in different cities or perhaps in different areas of the same city. One might question this approach immediately on the grounds that it requires preferences to be identical across hedonic markets. Accepting equality of preferences for the sake of argument, we investigate the conditions under which multiple market data will help solve the identification problem.

To keep the analysis simple, we suppose that the parameters of the hedonic models are estimated from other sources. Assume that we have G markets, and from each market we have a vector of hedonic parameters, γ^g, $g = 1,\ldots,G$ which we treat as known constants. We have two separate cases, depending on the functional form of the hedonic price equation. First we consider the case where all hedonic price equations are nonlinear.

4.4A Multiple Markets: Nonlinear Hedonic Prices

When $h(z;\gamma^g)$ is nonlinear, we cannot solve for the z's and are forced to work with the equilibrium conditions. Suppose, as before, that there are K characteristics. Then the equilibrium conditions for the g^{th} market for a household with income y^{jg}, attribute vector z^{jg} and hedonic price p^{jg} are:

$$
\begin{aligned}
h_1(z^{jg},\gamma^g) &= m_1(z^{jg},y^{jg} - p^{jg};\beta) + \mu_1 \\
\vdots \quad & \quad \vdots \\
h_K(z^{jg},\gamma^g) &= m_K(z^{jg},y^{jg} - p^{jg};\beta) + \mu_K
\end{aligned}
\tag{4.27}
$$

where, for convenience, we assume that $\mu \sim N(0,\Sigma)$. This model has K endogenous variables (z) and K structural equations, given in (4.27). Unless we make very restrictive assumptions about the utility function, the m_i functions will not be linear in parameters. Hence in general we can only establish the identifiability of the β through maximum likelihood estimation.

To get more insight into the multiple market setting, ignore temporarily the right hand side of (4.27). What is the role of the $h_i(z^{jg}, \gamma^g)$ here? It is clear that they are not exogenous, because they depend on endogenous variables, z^{jg}. They cannot be basic endogenous variables, because we already have equal numbers of equations and basic endogenous variables. In the language of Fisher or Goldfeld and Quandt, the h_i are nonbasic or additional endogenous variables. Unless attributes enter the hedonic price equation identically, each marginal price function (h_i) will be different. Thus there are K nonbasic endogenous variables, one in each equation. Since h_i enters only the i^{th} equation, each equation has K-1 excluded additional endogenous variables.

These K-1 restrictions for each equation are clearly of value, whether the model is linear in parameters or not. In the simplest linear-in-parameters case, K restrictions are needed to identify the i^{th} equation (cf. equation (4.11) where M = K), so that only one restriction would be required. The restrictions are also of value in the nonlinear-in-parameters case, where ML methods are necessary, since restrictions help ensure the nonsingularity of the information matrix.

Some caution must be exercised in the interpretation of these results, especially with regard to G, the number of markets available. G must be large enough to provide independent variation in the h_i. The reasoning for identification of linear-in-parameters, nonlinear-in-variables requires that the additional endogenous variables be asymptotically uncorrelated with endogenous variables. The same kind of argument would hold in this case. The h_i must be asymptotically uncorrelated with the z. Thus it is not the existence of two or more markets which guarantees identification, but the existence of enough markets to ensure some orthogonality between the h_i and z_i.

4.4B Multiple Markets: Linear Hedonic Prices

The utility of multiple markets is greatly enhanced by linearity in the hedonic price equation. The estimation problem in (4.27) can be transformed to a standard demand system when $h(z;\gamma)$ is additive and linear in z. In that case the marginal prices are

$$h_i(z^{jg},\gamma^g) = \gamma_i^g$$

and the system becomes (ignoring the μ_j)

$$\begin{aligned} \gamma_1^g &= m_1(z^{jg},y^{jg};\beta) \\ &\vdots \\ \gamma_K^g &= m_K(z^{jg},y^{jg};\beta). \end{aligned}$$

We can then solve for z^{jg} as in (4.5) and (4.6):

$$z^{jg} = D(y^{jg},\gamma^g;\beta) \qquad (4.28)$$

where now the γ's play the role of prices in linear budget constraints. If there are enough markets, then the variation in γ^g, being exogenous to the individual household's behavior, will allow the estimation of a demand system. The best example of this approach is provided by Parsons (1985) who estimates the almost ideal demand system for attributes using multiple city data. As in other situations, we can make tradeoffs between price information and the complexity of the model we estimate. For example, if we make the preference function additive, we need variation in only one of the γ_i^g. Further variation in relative prices can be gained from the requirement that equation (4.28) be homogeneous of degree zero in y^{jg} and γ^g.

Thus we see that multiple market data definitely aids in identifying parameters of the preference functions. It can do so only by maintaining a specific hypothesis about the preference structure - that it not include the hedonic price as an argument — which in turn allows the testing of the necessary result that the hedonic price equation be linear.

4.5 Conclusion

This chapter has addressed the general problem of the identification of the parameters of hedonic models. Three basic questions were addressed:

1. When we estimate a hedonic system, what are we seeking? It was shown that the so-called hedonic demand function is really a marginal rate of substitution function embodying the parameters of the preference function. As long as the hedonic price equation is nonlinear, traditional direct or inverse Marshallian and Hicksian demand functions do not exist as solutions to the consumer's choice problem. Estimation of a hedonic system is therefore an attempt to recover the consumer's preference parameters.

2. Under what circumstances is it possible to identify the preference parameters? Necessary and sufficient conditions were derived for the identification of the parameters of recursive and nonrecursive single-market hedonic models and multiple market models. Models linear in parameters and models nonlinear were investigated. As with all econometric identification problems, identification is dependent on prior restrictions imposed on the parameters and functional form of the equations in the model. Unlike the traditional problem of identifying supply and demand functions by exclusion of variables, very few theory-based restrictions are available for hedonic models. Identification instead requires the imposition of generally untestable restrictions on the functional form of the hedonic and marginal rate of substitution equations. These restrictions often place unknown or unrealistic limitations on the underlying preference or market structure. Our results on identification my be summarized by the following:

i) Identification must be determined by prior considerations. In particular, there are no circumstances where one can apply the Rosen two-step approach without imposing prior constraints and be assured of identification.

ii) Successful estimation of a hedonic system by maximum likelihood techniques is sufficient to demonstrate the existence of an identified model.

iii) When the parameters of the hedonic price equation are known (available from another source) it may be possible to solve for the attributes' reduced form equation. The system will then be identified if it is possible to derive the preference parameters from the reduced form estimates.

iv) The use of data from multiple markets definitely aids in the identification of the preference parameters, though it is still necessary to impose severe restrictions on the underlying preference structure.

v) The conditions for identification just discussed are technical, relating to the application of traditional criteria to the rather special case presented by hedonic markets. But the fundamental question of identification relates to behavior: What kind of behavior must we assume to achieve identification and are we likely to find such behavior in the real world? The answers to this compound question are not very satisfying, mostly due to the nature of the hedonic price equation. This equation, which is structural to the household, reflects the combined influence of buyers, sellers and the distribution of goods. Restrictions on the functional form of the hedonic price equation may help satisfy the technical criteria, but restrictions cannot be translated into information about the behavior of buyers and sellers. As we show in Chapter 6, characteristics of buyers and sellers are likely to be masked in the hedonic equation. Of course, we also need restrictions on the marginal rate of substitution functions. The restrictions which are most likely to be useful are separability restrictions on the utility function. For example, the elementary rule of having the number of excluded exogenous variables exceed the number of included endogenous variables is helped by separability, because it means fewer endogenous variables in each marginal rate of substitution equation. There are few tests of separability in the hedonic setting, but it seems a safe bet that real world behavior does not support much separability. In sum, we can describe behavior needed to support identification, but we cannot find strong arguments to support the common practice of such behavior.

3. Is the solution to the identification problem worth the restrictions we must impose? The cost of identification come in the form of maintaining very specific and restrictive hypotheses about preferences and the hedonic price equation. The restrictions required for identification in the hedonic model are especially disturbing because they involve functional form rather than the exclusion of exogenous variables. Thus they lack the intuitive appeal of the more traditional approaches to identification. For example, in supply and demand models of agricultural commodities, we can identify demand by excluding rainfall from the demand function. No such appealing restrictions appear to be available in hedonic models. The benefits of recovering the parameters depend on how they will be used and whether in fact the hedonic model is suitable for valuing environmental amenities. In succeeding chapters, we show that there is a number of serious problems in using the hedonic model for measuring welfare effects, even when all parameters are known perfectly. We will thus postpone until the concluding chapter a full response to the questions of whether the solution is worth the cost.

CHAPTER 4

FOOTNOTES

1 McConnell is with the Department of Agricultural and Resource Economics, University of Maryland, and Phipps is with Resources for the Future.

2 And thus the exchange between Pollak and Wachter and Barnett is especially relevant.

3 Properties of inverse demand functions are derived from the problem

$$\min_{p}\{V(p,y) \mid px - y = 0\} \qquad \text{(i)}$$

where

$$V(p,y) = \max_{x}\{U(x) \mid px - y = 0\}$$

and where x and p are the vectors of goods and prices respectively and $V(\cdot)$ and $U(\cdot)$ are respectively the indirect and direct utility functions. Suppose the nonlinear budget constraint is $h(x,\gamma) - y = 0$, where γ is a vector of parameters. Then the indirect utility function becomes

$$V(\gamma,y) = \max_{x}\{U(x) \mid h(x;\gamma) - y = 0\} \ .$$

But there is no well-defined dual such as (i) which yields the inverse demand functions in this case.

4 For the motivation of these criteria, see Fisher (1966), Chapter 6, and Goldfeld and Quandt (1972) Chapter 8. They are analogous to the conditions for linear-in-variables systems when the additional endogenous variables play the role of exogenous variables.

5 The regularity assumption requires that the information matrix be of constant rank in an open neighborhood of θ_0.

1. A Cobb-Douglas Example

Consider the following example. Let the attribute vector be one-dimensional. For simplicity, let preferences be given by the Stone-Geary function:

$$U = \beta_1 \ln z + \beta_2 \ln(x - \beta_3) \tag{4.A.1}$$

and suppose that the hedonic price equation is given by

$$h(z;\gamma) = \gamma_0 z^{\gamma_1}$$

so that the budget constraint is given by

$$y = \gamma_0 z^{\gamma_1} + x.$$

The goal is to solve for the choice variables z and x. The equilibrium conditions are

$$h(z) = \gamma_0 z^{\gamma_1}$$

$$\gamma_0 \gamma_1 z^{(\gamma_1 - 1)} = \frac{\beta_1}{\beta_2} (y - \gamma_0 z^{\gamma_1} - \beta_3)/z.$$

Solving the equilibrium conditions for z and x gives the demand functions D and D_x analogous to (4.5) and (4.6):

$$z = (\beta_1/(\beta_1 + \beta_2\gamma_1))^{\frac{1}{\gamma_1}} \gamma_0^{-\frac{1}{\gamma_1}} (y - \beta_3)^{\frac{1}{\gamma_1}} \tag{4.A.2}$$

$$x = \beta_1\beta_3/(\beta_1 + \beta_2\gamma_1) + y(1 - \beta_1/(\beta_1 + \beta_2\gamma_1)). \tag{4.A.3}$$

These are demand functions in the sense that only exogenous variables are on the right hand side. But they are not traditional because neither marginal nor average price appears on the right hand side. The demand function collapses to the traditional Marshallian demand function when $\gamma_1 = 1$, implying that the hedonic price function is linear. This case is a linear expenditure system demand function because of the form of the preference function in

(4.A.1):

$$z = \beta_1(y - \beta_3)/\gamma_0 \qquad (4.A.4)$$

where it is assumed without loss of generality that $\beta_1 + \beta_2 = 1$. This is of course just the demand function in the linear expenditure system with a zero level of the subsistence parameter for z. The expressions for z and x in (4.A.2) and (4.A.3) can be written:

$$z = \pi_{01}(y - \pi_{11})^{\pi_{21}} \qquad (4.A.5)$$

$$x = \pi_{02} + \pi_{12}y \qquad (4.A.6)$$

when

$$\pi_{01} = (\beta_1/(\beta_1 + \beta_2\gamma_1))^{\frac{1}{\gamma_1}} \gamma_0^{-\frac{1}{\gamma_1}} \qquad (4.A.7)$$

$$\pi_{11} = \beta_3 \qquad (4.A.8)$$

$$\pi_{21} = 1/\gamma_1 \qquad (4.A.9)$$

$$\pi_{02} = \beta_1\beta_3/(\beta_1 + \beta_2\gamma_1) \qquad (4.A.10)$$

$$\pi_{12} = 1 - \beta_1/(\beta_1 + \beta_2\gamma_1) \, . \qquad (4.A.11)$$

Although there are five reduced form parameters and five structural parameters, we cannot solve for the γ's and β's without more prior information. Note that $\pi_{02}/\pi_{11} + \pi_{12} = 0$ for all values of β_1, β_2 and γ_1, and hence there is a redundancy. However, by imposing the prior constraint $\beta_1 + \beta_2 = 1$ we can solve for the β's and γ's and hence solve the identification problem.

In the case where the γ's are known with certainty (section 4.6), we have reduced form equations for attributes only. Then we estimate (4.A.5), imposing the prior constraint (4.A.9). This leaves two reduced form parameters, π_{01} and π_{11} and two structural parameters (assuming $\beta_1 + \beta_2 = 1$): β_1 and β_3. Since (4.A.8) tells us where to get β_3, we need only solve (4.A.7) for β_1.

Imposing $\beta_1 + \beta_2 = 1$ and solving (4.A.7) for β_1 yields

$$\beta_1 = -\gamma_1/(1 + \gamma_1 - \gamma_0\pi_{01}^{-\gamma_1}).$$

For this particular hedonic equation, knowing the γ's simply reduces the estimating problem, with no fundamental change in the identifiability of the β's.

This one attribute example shows the difficulty of solving for the demands. Irish (1980) has developed cases which can be solved, but the necessary simplifications show the difficulties involved.

2. A CES example

A separate example illustrates the contention that simply assuming a utility function and applying the Rosen two-step approach is no guarantee of identification.

Suppose the hedonic price equation is as before but that the preference function is given by the GCES:

$$U(z,x) = \beta_1 z^{\beta_2} + \beta_3 x^{\beta_4} .$$

The equilibrium condition is

$$\gamma_0 \gamma_1 z^{(\gamma_1 - 1)} = (\beta_1 \beta_2 / \beta_3 \beta_4) z^{(\beta_2 - 1)} (y - p)^{(1-\beta_4)}. \qquad (4.A.12)$$

We can use the Rosen two step on this expression (in logarithms, as in Quigley, 1982) with errors added on. The model to be estimated from the logarithm of (4.A.12) is

$$h_i = \delta_0 + \delta_1 \ln z + \delta_2 \ln(y - p) + \text{error} \qquad (4.A.13)$$

when $\quad h_i = \ln(\gamma_0 \gamma_1) + (\gamma_1 - 1) \ln z$

$$\delta_0 = \ln \beta_1 \beta_2 / \beta_3 \beta_4$$

$$\delta_1 = (\beta_2 - 1)$$

$$\delta_2 = (1 - \beta_4)$$

are parameters to be estimated. An application of OLS to (4.A.13) yields

$$\hat{\delta}_0 = \ln(\gamma_0 \gamma_1)$$

$$\hat{\delta}_1 = (\gamma_1 - 1)$$

$$\hat{\delta}_2 = 0.$$

We simply reproduce the parameters of the hedonic price equation as in the examples given by Brown and Rosen. Hence even though we recognize that we should be estimating the marginal rate of substitution conditions, we still have ample room to create a constructed marginal price problem.

The subtle nature of the constructed marginal price problem can be illustrated if we impose the restriction that preferences are homothetic, so that the utility function becomes the CES. The logarithm of the equilibrium condition becomes, on imposing $\beta_2 = \beta_4 = \beta$,

$$h_1 = \delta_0 + \delta_1 \ln(z/(y - P)) + \text{error} \qquad (4.A.14)$$

where $\delta_0 = \ln(\beta_1/\beta_3)$ and $\delta_1 = (\beta - 1)$. Applying OLS to (4.A.14) does not imply that the estimates of • repeat the parameters of the hedonic price equation even when the power function for $h(z;\gamma)$ is used. While this example is perhaps too simple to consider for applications of the hedonic method, it illustrates the difficulties of hypothesis testing in this approach. For the GCES preference function, given the power function for the hedonic equation, the constructed marginal price problem makes the structural estimation meaningless. But when the CES preference function is imposed there is no longer a constructed marginal price problem.

CHAPTER 5

THE STRUCTURE OF PREFERENCES AND ESTIMATION OF THE HEDONIC PRICE EQUATION

K. E. McConnell and T. T. Phipps[1]

5.1 Introduction

In the previous chapter, we tried to determine the circumstances under which it is possible to identify the parameters of preference functions. In Chapter 2, we surveyed the practical problems encountered in using ordinary least squares on the hedonic price equation. The joint problems of multicollinearity, errors in specification and functional form plague the single equation estimates of hedonic price equations in housing markets (see Bartik and Smith, 1984 and Palmquist, 1983 for additional details). The issues which have arisen in estimating the hedonic price equation are primarily of a measurement nature, having little to do with simultaneity.

In the chapter 4, we developed the nature of the choice problem for the household. We argued that in an econometric structure which models the choice of the attributes and the price of the commodity, it makes sense to designate these same variables as endogenous. Then the hedonic equation is a part of the structural equation: the household's nonlinear budget. If the hedonic equation is in fact structural to the household, then it must be subject to possible under-identification. In this chapter, we follow the logic of Chapter 4 to investigate the circumstances under which the hedonic equation will be identified. These circumstances relate to the structure of preferences.

In this chapter, we will first show that the hedonic price equation may reasonably be considered part of the simultaneous system, then derive the circumstances when the hedonic equation can be consistently estimated with OLS, and finally, develop some Monte Carlo results showing the effects of simultaneity on OLS estimates of the parameters of the hedonic price equation. [illegible] applied to data on prices and attributes collected from market transactions. Hedonic price equations fitted on housing prices which are household's own estimates will obviously not be subject to any simultaneous equation issues because such estimates will not have been jointly determined with the purchase of attribute levels.

This chapter has two rather different purposes. First, it is designed to explore simultaneity in hedonic markets by developing the logical consequences of this simultaneity for the hedonic price equation. This chapter is not designed to critique the practice of estimating hedonic price equations. It would be foolhardy to assert that, in the midst of such pressing data problems and with so many attributes, one should worry about identification. Rather, we are trying to learn about choice in hedonic markets. The second purpose is more practical. Multicollinearity is a serious problem in hedonic models. But in nonlinear systems, the distinction between multicollinearity and under-identification is blurred. We argue that what is apparently multicollinearity may be endemic to the system precisely because of underidentification. In that case, the cure for multicollinearity of enlarging the sample size may simply cause parameter estimates to converge on the wrong values.

One conclusion of this chapter relates to the requirements for successful estimation of the parameters of the hedonic price equation. For analyses using market transactions, it will be shown that consistent estimates of these parameters require the assumption of restrictions on the form of the utility function. These restrictions will, in general, be untestable. This conclusion is quite similar in spirit to the received literature on identification of the parameters of preferences. In concluding their paper on identifying parameters relating to preferences, Diamond and Smith (1985) note

> Consistent estimation of the structural parameters of demand requires sufficient restrictions to identify functions. The minimum requirements can be met through the assumption of a utility function and hedonic function which imply the presence in the marginal price function of appropriate nonlinear transformations of the endogenous variables in the demand function. However, this approach relies heavily on the choice of utility function, while providing no independent statistical means to test that choice (p. 281).

We will argue that consistent estimation of the hedonic price equation by ordinary least squares with market transactions also requires making assumptions about the functional form of the hedonic price equation and the preference function.

5.2 The Structure of Preferences and the Equilibrium Conditions

In Chapter 4, the following choice problem was described for the household (section 4.3)

$$\max_{x,z}[U(x,z;\beta) \mid y-h(z;\gamma) - x = 0].$$

When there are K attributes, this problem has K+2 first order conditions:

$$U_x = \lambda \tag{5.1}$$

$$U_i = \lambda \partial h/\partial z_i \qquad i = 1,K \tag{5.2}$$

$$y - h(z;\gamma) - x = 0 \tag{5.3}$$

where λ is the multiplier on the income constraint and $U_i \equiv \partial U/\partial z_i$. These first order conditions yield solutions for the K+2 variables (x,λ,z) of the optimization problem. By the substitutions of 4.3, this system can be reduced to K+1 equations in the K+1 variables natural to the consumer (p,z):

$$p = h(z;\gamma)$$

$$\partial h(z;\gamma)/\partial z_i = m_i(y - p,z;\beta) \qquad i = 1,K$$

where, as before $m_i(y - p,z;\beta) \equiv \partial U(x,z;\beta)/\partial z_i/\partial U(x,z,\beta)/\partial x$ evaluated at $x = y-p$. We give this system an additive error structure which we consider to be econometrician's error and which captures the spirit of empirical efforts in hedonic modelling. Then our system is

$$p = h(z;\gamma) + \varepsilon_1 \tag{5.4}$$

$$\partial h(z;\gamma)/\partial z_i = m_i(y - p,z;\beta) + \varepsilon_{2i} \quad i = 1,K. \tag{5.5}$$

The purpose of deriving (5.4) and (5.5) is to make clear the origin of the system. It is a structural representation of the household's optimization for a nonlinear budget constraint. In general, hedonic models are concerned with recovering the parameter vectors γ and β. The discussion of the identification problem has focused on the difficulties of estimating the β's and how they can be confused with the γ's, as for example, Brown and Rosen (1982) have shown. However, we can also see that, it is possible in principal to confuse the γ's with the β's. Our focus here will be on the problems of recovering the parameters of the hedonic price equation. Specifically, how do the values of β influence the identifiability of the γ's?

5.3 Estimation of the Hedonic Price Equation

In this section we ask under what conditions we can estimate the parameters γ using single equation methods. While there have been numerous efforts to use Box-Cox techniques (for example, Halvorsen and Pollakowski, 1981), we will assume linear-in-parameters models. Nonlinearity would complicate the form but not alter the substance of the argument.

Let equation (5.4) be written as linear-in-parameters:

$$p = h(z)\gamma + \varepsilon_1$$

where $h(z)$ is a vector of functions of the z's and $h(z)$ and γ are conforming vectors of dimension J, where J is less than the number of observations. The OLS estimates of γ are

$$\hat{\gamma} = \gamma + (h^T h)^{-1} h^T \varepsilon_1. \tag{5.6}$$

Note that h, being a function of z's, depends on ε_2. Further, if p is in m, then the z's depend on ε_1. Hence, h is a random function of ε_2 and possibly ε_1. The randomness of h and the nonlinearity of random terms in $(h^T h)^{-1} h^T \varepsilon_1$ make it difficult to give general statements about the bias in γ. But we know that the consistency of $\hat{\gamma}$ requires that

$$\text{plim } h^T \varepsilon_1 = 0. \tag{5.7}$$

For expression (5.7) to hold, we must have the vector of z's uncorrelated in the limit or distributed independently of ε_1. Since expression (5.5) can be solved for z_i, $i = 1,K$ in principle, we could have z as a function of p and ε_2, or substituting for p, have z depending on ε_1 and ε_2. Thus we see in general that (5.7) will not hold, so we need to look closer at what assumptions will make it hold.

Suppose first that m is independent of p. Then z, and hence h, are functions of ε_2 only. Blips in ε_2 will influence h, but h will move systematically with ε_1 only when ε_1 and ε_2 are correlated. Hence, correlation between ε_1 and ε_2 will cause inconsistency. Now if m depends on p, the solution for z depends on ε_1 and ε_2, causing $\hat{\gamma}$ to be inconsistent. Thus we have two requirements for consistency of $\hat{\gamma}$:

1. $m_i(y - p, z; \beta)$	independent of p;
2. ε_1 and ε_{2i}	uncorrelated in the limit.

These requirements, which must hold for all i, of course, are simply the requirements for recursivity in nonlinear systems. But how restrictive are they?

Consider first the requirement that m_i not depend on p. Let us consider a particular m_i:

$$0 = \frac{\partial m_i}{\partial p}$$

$$= \frac{\partial m_i \partial x}{\partial x \partial p}$$

$$= - \frac{\partial m_i}{\partial x}$$

because $x = y - p$ and $\frac{\partial x}{\partial p} = -1 \neq 0$. Recalling that $m_i(y - p, z; \beta) = m_i(x, z; \beta) =$ $\partial U(x,z;\beta)/\partial z_i / \partial U(x,z;\beta)/\partial x$ we see that

$$0 = \frac{\partial m_i}{\partial x}$$

$$= \frac{U_{ix} U_x - U_i U_{xx}}{U_x^2} . \tag{5.8}$$

This condition must hold for $i = 1,K$ for consistency. The restriction that the numerator of (5.8) be zero is imposed on the preference function. It can be satisfied by the restriction $U_i/U_x = U_{ix}/U_{xx}$. Or it holds when U is linear in $\underset{\sim}{x}$. We can obtain it, for example, from the preference function $U(x,z) = x + \bar{U}(z)$ where $\bar{U}$ is any quasi-concave function of z. The assumption of m_i independent of p imposes restrictions on the preference function, restrictions which are as untestable as those needed to identify the marginal rate of substitution function through nonlinearity in the hedonic price equation.

The practical significance of the use of y rather than y-p may be tempered by the magnitude of y relative to p and by measurement errors in y. The relationship between income as measured in most survey work and income which constrains the household's budget must surely be prone to large errors. One cause of the difference, for example, would be real wealth holdings, which usually do not show up in current income figures. This would be especially important in home purchases. When coupled with large y relative to p, it seems intuitively plausible that such large errors would mask the omission of p from the argument y-p.

There is less reason to be reluctant to assume that ε_1 and ε_{i2} are uncorrelated. At least we have no reason to argue for correlation in one direction or another. But there is a strong tradition in demand systems analysis for correlation of errors across equations. Depending on the data source, one might argue for or against this correlation. Hence, it is the structure of the preference function which is the strongest requirement in obtaining consistent estimates of γ.

5.4 Some Monte Carlo Results on the Identifiability of the Hedonic Price Equation

To some extent, the question of whether the hedonic price equation is identified is an empirical one. That is, for some structures, the single equation estimates may be good enough. To test the degree to which OLS estimates miss the true value of hedonic parameters, we have done some simple Monte Carlo estimations for a model which we a priori know to be not identified. The model contains two attributes. The preference part of structure of the model is consistent with a linear approximation of the bid function. The hedonic equation is given by

[illegible]

and the equilibrium conditions are given by

$$\gamma_1 = \beta_{01} + \beta_{11}z_1 + \beta_{21}z_2 + \beta_{31}(y-p) + \varepsilon_{21} \quad (5.10)$$

$$\gamma_{12} + \gamma_{22}z_2 = \beta_{02} + \beta_{12}z_1 + \beta_{22}z_2 + \beta_{32}(y-p) + \varepsilon_{22}. \quad (5.11)$$

First we demonstrate using traditional criteria that the first equation is not identified. Let us write the system as in section 4.3 above:

$$Aq(w) = \varepsilon$$

where $q(w) = (p, z_1, z_2, 1, z_2^2/2, y)^T$ and

$$A = \begin{bmatrix} 1 & -\gamma_1 & -\gamma_{12} & -\gamma_0 & -\gamma_{22} & 0 \\ \beta_{31} & -\beta_{11} & -\beta_{21} & \gamma_1 - \beta_{01} & 0 & -\beta_{31} \\ \beta_{32} & -\beta_{12} & \gamma_{22} - \beta_{22} & \gamma_{12} - \beta_{02} & 0 & -\beta_{32} \end{bmatrix}.$$

Let Φ_1 be the matrix of restrictions on the parameters of the first equation. In section 4.3 (equations 4.9 - 4.12), it is argued that the necessary condition for identifying any parameter of the hedonic model is

$$\text{rank } (\Phi_i) \geq M - k_i$$

where M is the number of equations and basic endogenous variables in the model, and $k_i = 0$ (or 1) is the number of normalized parameters in the i^{th} equation.

In the case of the model above, $M = 3$ and $k_i = 1$, so the necessary condition for identification of the hedonic price equation is

$$\text{rank } (\Phi_1) \geq 2.$$

The only restriction placed on the equation is that y is excluded. Hence,

$$\Phi_1 = [\,0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1\,]^T$$

$$\text{rank } (\Phi_1) = 1 < 2.$$

Thus the necessary conditions for identifying the hedonic price equation are not met, and the equation is not identified according to the traditional criterion. Applying OLS to (5.9) will result in biased estimates of γ's. Further, as the sample size increases, the OLS estimates will not get closer to [illegible]

To demonstrate further with this example, we show the results of OLS applied to equation 5.9 using Monte Carlo methods. We have performed two different sampling experiments with the basic structure as given in (5.9) - (5.12). The experiments have the same distribution of the income variables and one of two possible distributions of errors. The income variable is drawn from a uniform distribution between 40 and 90. The errors are normally distributed errors with mean zero and diagonal covariance matrix:

$$\sum = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad (5.12)$$

or the nondiagonal covariance matrix:

$$\sum = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 2 \end{bmatrix}. \qquad (5.13)$$

The experiments use the coefficient matrix:

$$\begin{matrix} p & z_1 & z_2 & 1 & .5z_2^2 & y \end{matrix}$$
$$\begin{bmatrix} 1 & .9 & 3.5 & 15 & .5 & 0 \\ .075 & -.8 & .006 & 4 & 0 & .075 \\ .008 & -.5 & -.5 & 6.5 & 0 & .008 \end{bmatrix}.$$

The experiments consist of estimating the model 20 times for the 50 observations and 20 times for the 500 observations. Various measures of the performance of estimators are given.

From Tables 5.1 and 5.2 we can get some feel (though not proof) of the properties of the estimates of γ. Consider first the diagonal error covariance case (Table 5.1). The relative bias of the $\hat{\gamma}_0$ is small and gets smaller as the sample size increases. The relative bias of $\hat{\gamma}_1$ is also small but shows only a barely perceptible change with the increase in sample size. The bias in $\hat{\gamma}_{12}$ grows with sample size. The bias in $\hat{\gamma}_{22}$ is uncomfortably large, but decreases marginally with the increase in sample size. When we consider the nondiagonal error covariance case (Table 5.2), we find that the $\hat{\gamma}_1$, and $\hat{\gamma}_{22}$ have bigger relative biases with higher sample size. For $\hat{\gamma}_{12}$ the bias improves, though the relative error is eight percent. In both cases the relative bias of $\hat{\gamma}_{22}$ appears substantial.

Of course, these results simply confirm what theory tells us, but they do also add some concreteness to theory. The basic result is that we cannot be absolutely confident that when we regress the transactions prices on the attributes of the good that we will recover the parameters of the hedonic price equation even when the data are [illegible]

TABLE 5.1

Monte Carlo Results for Hedonic Parameters
Diagonal Covariance Matrix
5.12

Parameters		Expected Value	Variance	Bias	Relative Bias	Mean Squared Error
$\hat{\gamma}_0$	T[a]= 50	15.448	1.705	.448	.0299	1.906
	T = 500	15.298	.2352	.298	.0199	.3240
$\hat{\gamma}_{11}$	T = 50	.8397	.0264	-.0603	-.0670	.0300
	T = 500	.8392	.0033	-.0608	-.0675	.0070
$\hat{\gamma}_{12}$	T = 50	3.833	.6302	.333	.0951	.7411
	T = 500	3.855	.0663	.355	.1014	.1923
$\hat{\gamma}_{22}$	T = 50	.0892	.1640	-.4108	-.8216	.3327
	T = 500	.1142	.0156	-.3858	-.7716	.1644

[a] T = sample size.

TABLE 5.2

Monte Carlo Results for Hedonic Parameters
Diagonal Covariance Matrix
5.13

Parameters		Expected Value	Variance	Bias	Relative Bias	Mean Squared Error
$\hat{\gamma}_0$	T[a]= 50	15.194	2.511	.194	.0129	2.549
	T = 500	15.15	.695	.150	.01	.717
$\hat{\gamma}_1$	T = 50	.8377	.8423	-.0623	-.0692	.8462
	T = 500	.8252	.0020	-.0748	-.0831	.0076
$\hat{\gamma}_2$	T = 50	3.905	.0686	.405	.1157	.2326
	T = 500	3.814	.0683	.314	.0897	.1669
$\hat{\gamma}_{22}$	T = 50	.1784	.1268	-.3216	-.6432	.2302
	T = 500	.1202	.0176	-.3798	-.7596	.1618

[a] T = sample size.

5.5 Conclusion

In the estimation of hedonic models from data on market transactions, the hedonic price equation and the marginal rate of substitution function form a simultaneous system. This chapter has undertaken to investigate the relationship between the structure of the marginal rate of substitution function and the consistency of OLS estimators of the hedonic price equation. Specifically we have shown that for consistent OLS estimators of the hedonic price parameters, the Hicksian bundle or income must not influence marginal values of attributes. This is a strong but generally untestable assumption which is not likely to hold in general.

Because most plausible preferences will violate the structure of recursivity, it may be that the parameters of the hedonic price equation are not identified. To test the nature of OLS estimates we performed some Monte Carlo experiments on several linear-in-parameters hedonic models. Our results showed that in some cases OLS estimators do not tend to get close to true values as the sample size grows.

Our results may provide some insight into the multicollinearity problem in the hedonic equations. Lack of identification shows up as perfect collinearity in linear and nonlinear two-stage least squares estimation. Further, as Wegge and Feldman (1983) have stated so succinctly, identification in nonlinear systems may sometimes be a matter of data and not structure:

> Instead of viewing the problem in a discontinuous fashion, one should perceive that the interface between identifiability, estimation, and prediction is a continuous relationship. Long before we reach the point of a discontinuous jump in the rank and its concomitant requirement of more prior information, we would be in a near singular moment matrix situation when the distinctions between some parameters become very confused, indicating that the parameter is close to not being identifiable (p. 253).

This description of the problem is quite apt for the hedonic price equation. Attributes which provide utility will tend to increase together with income and other socioeconomic measures. In this view, multicollinearity is simply a symptom of underidentification and may not be resolved as sample size increases.

CHAPTER 5

FOOTNOTES

1 McConnell is with the Department of Agricultural and Resource Economics, University of Maryland, and Phipps is with Resources for the Future.

2 Note that this derivative, and not the more complicated version imposing the first order conditions, is appropriate here.

CHAPTER 6

THE FORMATION AND USE OF THE HEDONIC PRICE EQUATION: A SIMULATION APPROACH

K. E. McConnell and T. T. Phipps[1]

6.1 Introduction

The purpose of this chapter is to look behind the veil of the hedonic price equation and into the workings of the market. To do so, we create a simulated market in which consumers choose housing locations, choosing attributes only implicitly because they are tied to locations. The market simulation allows us to explore two important issues in hedonic analysis: 1) the empirical connection between the parameters of the preference function and the hedonic price equation; and 2) the accuracy of four commonly used "restricted partial equilibrium" welfare measures (Bartik and Smith's phrase) in comparison to a true measure of welfare, given market adjustment. These two issues are closely related. Their resolution requires knowledge of the workings of the housing market: specifically, what is the nature of the equilibrium process which allocates households to sites? Further, welfare measurement directly or indirectly makes use of the hedonic price equation so the way this equation is estimated strongly influences welfare calculations.

The simulated market provides a good with three attributes. The supply of the good is fixed. For simplicity, the number of units of the good equals the number of buyers. The fixed supply is allocated to households as in a bid rent or utility maximization model. From this model, a price for each unit of the good is established. The price varies with the exogenously given attributes of the good, and hence is a hedonic price. In section 6.2a we describe the equilibrium of location choices. In section 6.2b we try to determine the effects of parameters of the preference function and different income distributions on the estimates of parameters of the hedonic price function. In section 6.3 we use the model to calculate partial and general equilibrium welfare effects of exogenous changes in the attributes of the fixed supply of goods.

6.2 Preferences, Income Distribution and the Functional Form of the Hedonic Price Equation

A component of current research in the implicit markets literature is that the structure of preferences is embodied in the hedonic price equation. One implication of this argument is that prior restrictions on the form of the hedonic equation may be derived from preference theory.

Rosen (1974) developed the theory in which the hedonic price function is generated by the competitive behavior of suppliers and demanders of goods containing a bundle of attributes $z = (z_1,\ldots,z_n)$. He argued that the hedonic function, $p(z)$,

> can sometimes be obtained if sufficient structure is imposed on the problem. However, it is not always possible to proceed in that manner. In general, the differential equation defining $p(z)$ is nonlinear and it may not be possible to find closed solutions. Moreover, a great deal of structure must be imposed. For example, the distribution of income follows no simple law throughout its range, making it difficult to specify the problem completely. Finally, partial differential equations must be solved when there is more than one characteristic (p. 48).

For these reasons, he recommended using the well known two step estimation approach in which the hedonic function is estimated first and then the calculated marginal prices are used to estimate what he calls the "marginal demand and supply functions."

Quigley (1982) used a simple fixed supply housing market example to demonstrate that the hedonic function may be derived by integrating the marginal rate of substitution function for a single hedonic attribute and a single Hicksian good. In his conceptual example, he assumed Cobb-Douglas preferences and the existence of a monotonic mapping from consumer income to the housing attribute.[3]

While Rosen and Quigley have demonstrated that the imposition of sufficient structure on preferences and income distribution (and supplier characteristics in the case of endogenous supply) in principle allows calculation of the hedonic price function, the empirical relationship between the structure of preferences and the form of the hedonic price function has not been explored. In this chapter, a simulation of an open city housing market, with given preference structure and a fixed supply of housing attributes, is used to examine this relationship. Two different utility functions (Stone-Geary and translog), and four different income distributions (uniform, segmented uniform, Pareto and normal) are used in the simulations. Box-Cox flexible forms are used in estimating the hedonic functions in each case. We find no clear empirical relationship between consumer preference parameters and the structure of the hedonic equation. Quite different mathematical structures are estimated for the hedonic [illegible] the distribution of income is varied, even with preferences and supply held constant. One implication of the chapter is that when the researcher is merely interested in estimating the hedonic function, use of a best fit approach, such as a Box-Cox flexible form, without taking account of consumer preferences, will, in the rare worst case, lead to a reduction in the efficiency of estimation. This case occurs only when we know the exact form of the hedonic price equation.

6.2a. The Allocation Model

In our model, consumers choose between locating within the city or on the periphery. The periphery is assumed to be composed of an undifferentiated agricultural plane. If all consumers have identical preferences and income, the existence of the "agricultural bundle," denoted z^A, available in unlimited quantities at a fixed price, p^A, sets an exogenous utility level, U^A, that may be used to solve for the equilibrium housing price structure. When individual incomes differ, U^A is still exogenous to households, but varies among households according to income. The bidding process will ensure that house prices within the city adjust such that all consumers achieve their exogenous utility levels set by z^A, p^A. This model is thus an open city model in the sense that household well being is fixed by exogenous factors. We have used this model because it makes the determination of equilibrium relatively simple. Note that this model requires only open competition for sites among buyers and sellers, with the potential for migration, to ensure equilibrium. No migration need occur. The real alternative is not migration but commuting.

The equilibrium in this model is determined by the adjustment of households. Households move among sites with exogenously given attributes until the households with highest incomes occupy the best sites, where best is determined by a separable component of the preference function. Having the preference function separable in z means that rankings among different bundles of z are not affected by other arguments of the utility function, in our case x or y - p. Thus if $U(x_0,z_1) > U(x_0,z_2)$ then $U(x,z_1) > U(x,z_2)$ for any x. Then the allocation can precede the determination of the hedonic price. Any ranking of sites based on the attributes will depend on the preference function. Different preference functions may give different rankings. Once household equilibrium is reached, the hedonic price is determined as if a monopolist owned the site. The hedonic price is bid up until each household, i, is just as well off as it would be with the agricultural bundle:

$$U(y_i - p_j, z_j) = U(y_i - p^A, z^A) \qquad (6.1)$$

where p_j is the hedonic price for bundle j. Expression (6.1) is the essence of the bid rent model.

The utility maximization model yields the marginal conditions which derive from the problem

$$\max_{x,z}\{U(x,z) \mid y = h(z) + x\}$$

or

$$\max_{z} U(y-h(z),z)$$

which of course yield the necessary conditions

$$h_i = U_i/U_x \tag{6.2}$$

where $h_i \equiv \partial h/\partial z_i$ and $U_i \equiv \partial U/\partial z_i$. Wheaton (1977) has demonstrated that at equilibrium, the "outer envelope of consumer bids exactly represents the price profile obtained when consumers maximize utility and demand balances existing supply" (p. 203). Hence, we may characterize the household equilibrium by the marginal bid functions:

$$h_i = B_i(z,y,U_i^A) \tag{6.3}$$

where the bid function B is defined implicitly in terms of utility by the function $U(y-B,z) = U^A$ and $B_i \equiv \partial B/\partial z_i$.

The structure needed to derive analytically or numerically the equilibrium hedonic price relation depends on the assumptions one is willing to make about how the market operates, consumer preferences and income distribution. For example, in the open city model, if all consumers have identical preferences and income, solution of the K partial differential equations given by (6.3), given the exogenous utility level, U^A, plus the boundary conditions, is sufficient to characterize completely the hedonic price equation.

In practice, the hedonic price equation is more complicated. It depends on the household allocation process as shown by Quigley (footnote 6, p. 183). Two characteristics of the preference function significant for the allocation process are:

1. Whether there is more than one attribute.
2. Whether the preference function is separable in the partition x and z.

If there is only one attribute, then the assignment of households to sites will be invariant to preferences. However, when there is more than one attribute, even simple preference functions of the same form but with different parameters will give different rankings. If the preference function is not separable in x,z, then the equilibrium allocation of sites to households depends on the equilibrium price. In order to know the rankings, one must know the hedonic price. With separability, the ranking is invariant to the hedonic price since the subutility function for housing attributes becomes "the quality hedonic rent function" (van Lierop, 1982, p. 281). In practice, the equilibrium hedonic price function would be solved numerically, and where there is no separability, iterative methods will be needed.

One noteworthy conclusion emerges concerning the open city model. Polinsky and Shavell (1976) have shown for a model with homogeneous households, "in a small open city the rent at any location depends on the level of amenities at that location" (p. 123). When incomes vary, this conclusion no longer holds. A change in the amenities at one site which changes the relative rankings of sites can cause a change in the whole hedonic gradient. This change occurs because even when the household's utility level is pegged by exogenous factors, as in the open city case, the assignment of households to sites must be done within the city. That is, we need some mechanism to describe equilibrium within the city when households are not identical. We have chosen the approach of allocating sites to the highest bidder. If the distribution of attributes among households changes in the sense that the rankings change, the equilibrium must also change. Imagine a change in the attribute vector that converts the worst site into the best. Then the rankings of all sites will change, and the price at each site must be recomputed. When households' preferences and incomes are the same, the assignment does not matter.

6.2b. Simulations

In this section we simulate the market described above, where preferences are identical, supply is fixed, and incomes vary. Our goal is to determine how hedonic price functions vary. The steps used in the simulations are:

i) Rank each housing bundle using the subutility function; then assign consumers to houses based on their income ranking. This is equivalent to assigning housing bundles to the highest bidder.
ii) Compute the exogenously determined utility level (U^A) each household would receive if the bundle z^A were bought at p^A.
iii) Calculate the price each household would have to pay for its respective site to give it the same utility level (U^A) it would receive if it chose the alternative bundle z^A at prices p^A.
iv) Estimate the hedonic equation, using a flexible functional form, by regressing a transform of the calculated prices on the hedonic characteristics.

Simulations were run using two different preference functions, the Stone-Geary:

$$U(x,z) = \beta_o \ln(x - \delta_o) + \sum_{j=1}^{3} \beta_j \ln(z_i - \delta_j) \qquad (6.4)$$

and the translog:

$$U(x,z) = \beta_o \ln x + \sum_{j=1}^{3} \beta_j \ln z_j + .5 \sum_{i}^{3} \sum_{j}^{3} \delta_{ij} \ln z_i \ln z_j . \qquad (6.5)$$

The parameters of the preference functions are [illegible] hundred and fifty house attribute vectors were generated using random drawings from the uniform distribution. Each vector contained three attributes. Four different distributions of income were generated: uniform, segmented uniform (a combined sample composed of drawings from two independent uniform distributions to simulate a segmented housing market), Pareto and normal. (The parameters of these income distributions are also in the appendix.) All incomes were scaled so that each distribution had a mean of 20,000. Hence, under any distribution of income, aggregate incomes are equal. Since both utility functions are separable, it was possible to rank each bundle using the housing sub-utility function. Housing bundles were then matched with incomes, and hedonic prices were calculated based on the bundle $z^A = (5,15,20)$ available at $p^A = 2000$.

One hedonic price function was estimated for each combination of preference functions and income distributions using Box-Cox flexible forms, similar to the approach of Halvorsen and Pollakowski (1981). The general form of the hedonic equation was:

$$\frac{p^{\theta} - 1}{\theta} = \sum_{i=1}^{3} \gamma_i z_i + \sum_{i}^{3} \sum_{j}^{3} \alpha_{ij} z_i z_j. \tag{6.6}$$

Table 6.1 gives values of θ for different models.[3] In general, the fits appeared excellent. T-statistics were very high and over 90% of the variation of the transformed dependent variable was explained.

While the estimated values of θ do not tell the whole story about functional form, they certainly play a big role. In these examples, the range of the estimates of θ is from -1.2 to .79. There are substantial differences in the behavior of the hedonic prices as a function of attributes.

TABLE 6.1

Transformation Parameter for Quadratic Box-Cox Hedonic Price Functions

Income Distribution	Preference Function	
	Stone-Geary	Translog
Uniform	.49	- .13
Segmented Uniform	- .47	- .87
Pareto	-1.2	-1.06
Normal	.79	.65

As is apparent from the above results, the parameters of the hedonic price function are sensitive to both the specific form of the preference functions and the distribution of income. While it is difficult to generalize, it seems that the hedonic function is more sensitive to variation in the distribution of income. For example, the maximum variation in θ, given the distribution of income, is .62 (uniform), whereas the maximum variation in θ, given the preference function, is 1.99 (Stone-Geary). This result is consistent with the presentation in the last section which showed that the hedonic function arises from the joint interaction of consumer preferences, income distribution, market structure and the characteristics of the existing stock of houses.

We conclude that our empirical ability to determine the influence of preference parameters on the hedonic price equation is virtually nil. For practical considerations, then, one may assume that the preference parameters and the parameters of the hedonic price function are not intertwined in any way that is not already obvious from examination of the consumer's equilibrium conditions. From the perspective of an empirical description of the housing market, when the desiderata are the parameters of the hedonic function, little will be lost by direct estimation of the hedonic equation, without taking preferences into account.

6.3 The Welfare Effects of an Exogenous Change in Attributes

We are ultimately interested in using the hedonic technique to determine the welfare effects of changes in air pollution and other environmental pollutants which influence the value of locations. Our simulation model provides a laboratory for experimenting with changes in exogenous attributes. By constructing the market, we can see precisely what happens as locations are improved.

Calculating welfare measures in hedonic markets raises a number of issues. These issues have been the focus of considerable and deserved attention. Work by Freeman has been especially crucial here (especially 1971, 1974a and 1974b); in addition, papers by Polinsky and Shavell (1976); Polinsky and Rubinfeld (1977); Scotchmer and Fisher (1980); Bartik and Smith (1984) and Brookshire *et al.* (1982) have dealt with the problem.

In this section, we appraise five welfare measures using the market that we have created. The advantages of this [illegible] calculations before and after adjustment to an exogenous change in environmental quality.

In the following section, we investigate the welfare effects on a change in z_1. Using this attribute as an instrument requires some explanation because z_1 is, after all, an endogenous variable in all the models of attribute choice so far investigated. However, z_1 is exogenous at the aggregate or market level, since its physical distribution cannot be influenced by household behavior. We can imagine the following events. A government agency institutes a policy which improves air quality. With households remaining at their houses, this change in air quality is exogenous. Under a variety of circumstances, however, the change in this attribute will disturb households' locational equilibrium. Households will then relocate according to the equilibrium mechanism, and at the new equilibrium, according to hedonic theory, prices will appear 'as if' households chose attribute levels.

Initially we calculate five kinds of welfare effects. The first four are estimates of the benefits of an increase in z_1 assuming that no relocation occurs (partial analysis). The fifth is the change in the hedonic price at the site after relocation and a new equilibrium is established. The five measures are:

M1: Suppose we have solved the identification problem, so that we have the parameters of the marginal rate of substitution function. Then we can compute the change in the area under the marginal rate of substitution schedule, holding the marginal utility of the numeraire constant. The marginal rate of substitution is given by

$$m_i(x,z) = \frac{\partial U(x,z)/\partial z_i}{\partial U(x,z)/\partial x} = \frac{\partial U(x,z)/\partial z_1}{\lambda}$$

where λ is the marginal utility of income and the price of x is unity. Holding λ constant, we have

$$M1 = \int_{z_1^o}^{z_1^*} m_1(x,z)\, dz = (U(x,z^*) - U(x,z^o))/\lambda. \qquad (6.7)$$

Note that M1 is in units of $\Delta\$ = \frac{\Delta\$}{\Delta U}\Delta U$. For λ approximately constant, M1 is approximately equal to the compensating variation for a change in z_1. Compensating variation, denoted CV, is defined by the expression

$$U(y - p - CV,\ z^*) = U(y - p,\ z^o). \qquad (6.8)$$

With λ constant, this expression can be written (via Taylor's series expansion because $\lambda = \frac{\partial U}{\partial x}$) as

$$U(y - p,\ z^*) - \lambda CV \doteq U(y - p,\ z^o)$$

Solving for CV gives

$$CV \doteq (U(y - p,\ z^*) - U(y - p,\ z^o))/\lambda$$

$$\doteq M1$$

when x is substituted for y - p. This measure is an exact measure of compensating variation only if the marginal utility of income is constant. M1 is typically the measure used when computing the area under a hedonic "demand" curve, as in Freeman's (1974a) equation (4). It requires that the identification problem be solved because the parameters of the utility function are needed.

The exact measure of compensating variation is calculated by solving equation (6.8) for CV, rather than solving the Taylor's series expansion. The result, (where U_y^{-1} denotes U inverted for y)

$$CV + p = y - U_y^{-1}[U(y-p,z^{\circ}),z^*]$$

is simply the household's bid for the house with attributes z*. The exact value for M1 is therefore

$$M1 = y - p - U_y^{-1}[U(y-p,z^{\circ}),z^*]. \qquad (6.9)$$

M2: The predicted change in the hedonic price:

$$M2 = \bar{h}(z^*,\gamma) - h(z^{\circ},\gamma) \qquad (6.10)$$

where γ is the vector of best parameter estimates for the hedonic price equation. This is an approach to computing the measure suggested by Lind (1973) as an upper bound approximation to the benefits of a public improvement. Brookshire <u>et</u> <u>al.</u> (1982) use M2 as an upper bound of the willingness to pay for improvements in air quality. This measure can of course be used without solving the identification problem.

M3: The linear approximation to M1:

$$M3 = \Delta z_1 \cdot \frac{\partial U/\partial z_1}{\lambda} \qquad (6.11)$$

where $\partial U/\partial z_i$ and λ are evaluated at the initial bundle. This measure is used by Harrison and Rubinfeld (1978) but since their "demand" function is linear, it amounts to the same as M1. Since in equilibrium $(\partial U/\partial z_1)/\lambda = \partial h(z,\gamma)/\partial z_1$, this measure is typically computed using the hedonic slope, thus not requiring that the identification problem be solved.

M4: The linear approximation to M2:

$$M4 = \Delta z_1 \cdot \partial h(z,\gamma)/\partial z_1. \qquad (6.12)$$

Since it is computed without the hedonic price equation M4 is a frequently used measure which may be viewed as an approximation of M1 or M2 because in the Rosen model the slope of the hedonic equation equals the "demand" function in equilibrium (see equation 6.2). Freeman (1974a) notes that in the standard equilibrium case this approximation will be biased upward (p. 81).

The measures M1 through M4 assume that the households do not move in response to the disequilibrium created by an exogenous change in attributes. The final measure, M5, is calculated after households move:

$$M5 = p^* - p, \quad (6.13)$$

where p^* is the price which emerges after relocation and p is the original price. This calculation was made from the actual prices at the locations. It accrues to landlords because, given the assumption of a small open city model, the utility of all homeowners will remain constant. Thus, the increase in rent, M5, is the maximum amount landlords are willing to pay rather than go without the change in the attribute. This measure is the correct one for the benefits of changing z_1 in this open city case, as stated by Polinsky and Shavell (1976): "In the open city, the change in the aggregate property values corresponds to the total willingness to pay on behalf of all parties" (p. 125). When aggregated across households, M5 correctly measures total benefits: "Benefits ... equal the total of all changes in land rents, positive and negative ..." (Lind, p. 189).

The computation of M5, the change in rent, requires the following steps:

i. compute $U_0(z^*)$, the separable part of the utility function, and rank the bundles according to $U_0(z^*)$;
ii. rank households according to their incomes;
iii. associate each household with the location of corresponding rank;
iv. calculate the hedonic price that would make the household indifferent between its equilibrium site and the opportunity bundle. This gives p^* from which M5 can be calculated.

For housing attribute improvements, M5 will exceed the exact measure of the restricted partial equilibrium welfare change, the maximum sum of households' bids for their current houses as given in (6.9). As long as only improvements occur, adjusting the equilibrium will allow some households to move to better houses, and none to worse houses. The open city assumption insures that each household's utility is constant, so that households will always pay their compensating variation.

The measures M1-M4 are calculated for each household experiencing a Δz_1 of 5 units, and summed across households for each distribution of [illegible] summed across sites for each distribution of income-utility function combination. These results are presented in Table 6.2.

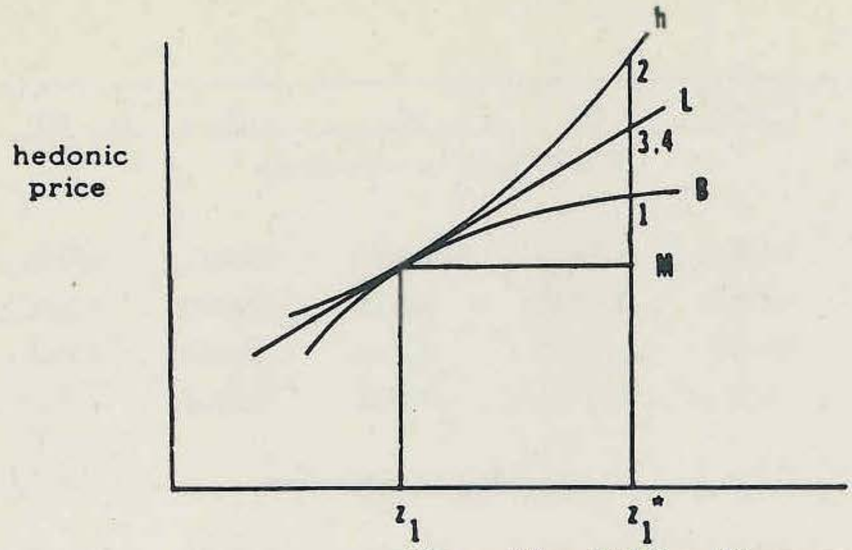
TABLE 6.2

Alternative Measures of Welfare for Exogenous Changes in an Attribute[a]
$\Delta z_1 = 5$

	M1[b]	M2	M3	M4	M5
			Stone-Geary preferences		
Income distribution					
uniform	85921	219540	121039	291012	85241
segmented uniform	74727	334641	99735	475685	74968
Pareto	64362	521563	85148	630229	63914
normal	77076	405051	103817	464901	77621
			Translog preferences		
Income distribution					
uniform	120162	193432	165925	319434	118450
segmented uniform	106154	89266	144197	287106	106523
Pareto	90109	8569	121251	11952	88922
normal	108316	355053	147425	474126	108985

[a] The initial range of supply is given in the appendix.
[b] The approximate measure calculated according to equation (6.7).

The calculations in Table 6.2 present some surprises which give insight not only into welfare measures but also into the working of the hedonic market. Order-of-magnitude errors are found in several different ways. M4 overstates M1 by almost an order of magnitude for the Pareto distribution and [illegible] magnitude. M2 and M4 typically overstate the other more acceptable measures. Let us look at the standard graphical analysis of M1, M2, M3 and M4 at equilibrium. Figure 6.1 shows the equilibrium as the tangency between h(z) and the bid function at z_1:



Alternative Welfare Measures
Figure 6.1

The best measure of the value of an increase in z_1, assuming no relocation, is M1 in Figure 6.1. At equilibrium, the marginal rate of substitution between the numeraire and z_1 equals the slope of the hedonic price equation. Hence M4 should equal M3, and with concavity of the bid function,

$$M4 = M3 > M1. \tag{6.14}$$

Further, when h(z) is convex, we have

$$M2 > M3 = M4. \tag{6.15}$$

Now let us look at Table (6.2). We find the following observations:

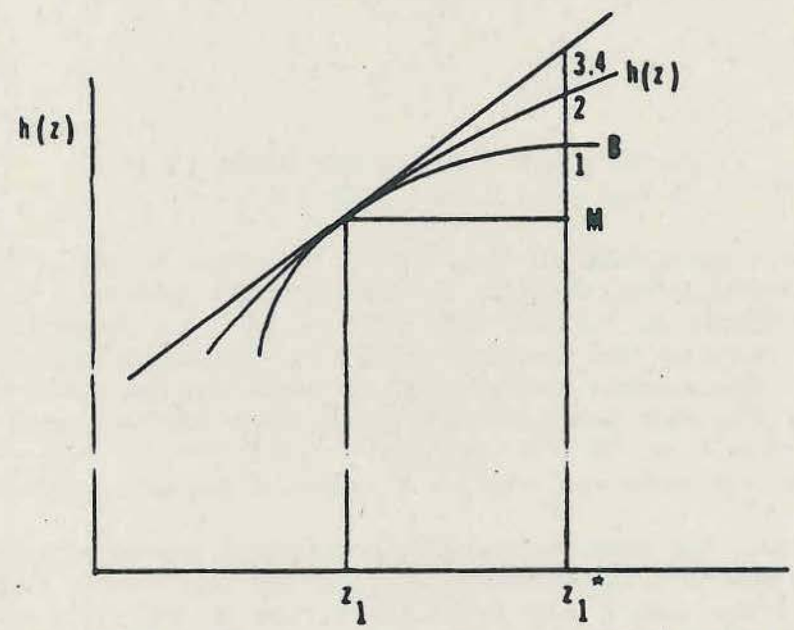
a. M4 > M3 implies $\partial h/\partial z_1 > (\partial U/\partial z_1)/\lambda$

b. M4 > M2 implies $\Delta z_1 \partial h/\partial z_1 > h(z+\Delta z) - h(z)$

The result (a) violates the idea that each attribute is in equilibrium at the margin. Result (b) contradicts the convexity of the hedonic price equation.

These results shed some light on the hedonic practices. They pertain primarily to the use of the hedonic price equation. Consider (a). We know that the equilibrium process ensures that at the margin, each household bids its willingness to pay for the i^{th} attribute. Yet when we complete the process of estimating the equilibrium bids as functions of attributes, and calculating the marginal hedonic prices, we find considerable differences between the known marginal bid and the slope of the hedonic price function. There are two explanations for these differences. First, the number of households is finite, and we have only points on the hedonic price function, not the exact function. Second, while all hedonic functions fit well, they still fit imperfectly, and the nonlinearity of the hedonic slope will in general prevent its expectation from equaling its true value. That is, the expectation of a function of a random variable will typically not equal the function of the expectation of the random variable, except when the function is a simple linear one.

Result (b) suggests that we could draw the hedonic price equation as in Figure 6.2. This shows the hedonic price equation to be concave in the area of some z_1's. First, this does not violate optimality conditions because they require only that h(z) be less concave than the bid function. Second, from a practical econometric perspective, nothing about the choice of functional form of the hedonic price equation restricts the chosen function to having the right curvature. Thus, while the Box-Cox method may allow the researcher statistical flexibility, it makes it harder to keep track of whether the apparent household equilibria fulfill the appropriate convexity conditions.



Welfare Measures in a Concave Neighborhood for h(z)

FIGURE 6.2

These results have implications for the identification problem. If we can calculate acceptable benefit measures from the slope of the bid function (M3), and we are confident that the households' equilibrium at the margin holds (M3=M4), then we could neglect the identification problem. From Table 6.2, we can see that linear extrapolations of the marginal bid (M3) provide 'in the ball park' approximations of M1. M3 exceeds M1 by 35% - 40%. This result depends on the parameters of the preference function, and cannot be generalized. But what is more important is that the hedonic slope misses the marginal bid considerably. Since in applications, our only knowledge about marginal bids comes from the slope of hedonic equations, it would seem somewhat premature to worry about the identification problem. Consequently, one conclusion from this simulation is that we need to know more about the distribution of the slopes of the hedonic price equation.

As a consequence of the discrepancies in welfare measures, we have discarded M2, M3, and M4 for further experiments and will concentrate on the restricted partial equilibrium measure of willingness to pay (M1) and the actual change in rents (M5).

Table 6.2 shows that the change in rents after the relocation is quite close to the households' approximate willingness to pay in the restricted case. In order to assess the potential magnitude of differences we have calculated M1 and M5 for three additional changes in z_1:

i) $\Delta z_1 = 1$

ii) $\Delta z_1 = .2 z_1$

iii) $\Delta z_1 = \begin{cases} 8 \text{ for worst half of the sites (1-125)} \\ 0 \text{ for other sites (126-250).} \end{cases}$

These results are presented in Table 6.3. In cases (i) and (ii) there is little change in the equilibrium because all bundles are improved, and little reason to expect differences in M1 and M5. Hence we have approximated M1 as in equation (6.7), keeping the marginal utility of income constant. In case (iii), where there is considerable reshuffling, we calculate the exact M1 according to equation (6.9). The two measures are quite close for the small changes in (i) and (ii). For case (iii), the change in the restricted willingness to pay tends to be 2% to 10% less than the change in rents, a result consistent with theory.

Finally, recall that mean household income and hence aggregate income are the same in all models. Consequently, given the preference function, the only reason for variation among the measures is the distribution of incomes. For the case of substantial distributional change in the attractiveness of the sites (iii), there is more than a two-fold difference in the extremes of the estimates of changes in rents. This case occurs when we compare M5 for the uniform (62180) and Pareto (30545) distributions of income. This result is one of aggregation and while the qualitative aspect is not surprising, the size of difference is. It suggests that the distribution of income is an important determinant of willingness to pay for changes in air quality, and that substantial inaccuracies can occur by ignoring this distribution.

TABLE 6.3

Further Comparisons of Welfare Changes

	M1			M5		
Δz_1	1[a]	$.2z_1$[a]	$\left\{\begin{smallmatrix}8\\0\end{smallmatrix}\right.$[b]	1	$.2z_1$	$\left\{\begin{smallmatrix}8\\0\end{smallmatrix}\right.$
			Stone-Geary preferences			
Income distribution						
uniform	21917	54356	59661	21879	53832	62180
segmented uniform	18415	50347	34412	18470	49940	39888
Pareto	15720	46914	29614	15688	46481	30545
normal	19112	51150	42971	19225	50778	47027
			Translog preferences			
Income distribution						
uniform	30847	72516	80176	30727	71243	84154
segmented uniform	26924	67218	48659	26949	66536	55889
Pareto	22687	60437	43281	22590	59711	44698
normal	27514	68064	58637	27612	67426	65247

a M1 is calculated according to (6.7), its approximate value.
b For this case, M1 is calculated according to (6.9), its exact value.

TABLE 6.4

Calculating Welfare as Changes in the Rent of Affected Sites Only

$$\Delta z_1 = \begin{cases} 8 \\ 0 \end{cases}$$

	Sum of Rent Changes (M5)	Sum of Rent Changes at Affected Sites Only
	Stone-Geary preferences	
Income distribution		
uniform	62180	48047
segmented uniform	39888	28441
Pareto	30545	24204
normal	47027	33706
	Translog preferences	
Income distribution		
uniform	84154	68956
segmented uniform	55829	42367
Pareto	44698	37327
normal	65247	50577

As our last experiment, we calculated what the estimate of benefits would be if, after relocation, we looked at the affected sites only. The only case where not all sites are affected is the case where $\Delta z_1 = 8$ for the worst half of the sites. We know from Freeman (1974b) and Lind (1973) that for this to serve as an upper bound, the willingness to pay must be identical among households. (This is directly related to the Polinsky-Shavell result that in a small open city, housing prices at any area location are independent of other [illegible] some rents go up and some rents go down when the equilibrium changes, but all households' willingness to pay will go up, because everyone moves to a better house. But in the open city case, we get the same result if we sum households' bids or landlords' rents, and we know that the sum of households' bids will increase if we allow adjustment. Therefore, looking at the rent changes at the affected sites only will understate the welfare change in the small open city when households differ by income. It is interesting to look at the magnitude of these rent changes and their variation across preferences and income distributions. The results are shown in Table 6.4, where the complete measure (sum of rent changes) is compared with the sum of rent changes on affected sites only. This table again shows the considerable variation in the measures across income distributions.

6.4 Conclusions

In this chapter we have simulated an open city housing market in order to investigate the determination of hedonic prices. This simulation market has allowed us to address two topics: (1) the influence of preference parameters and the distribution of income on the estimated functional form of the hedonic price equation and (2) the relationships among the various restricted measures of welfare and the post-adjustment change in rent, all induced by an improvement in the attributes of locations.

There are two principle findings with regard to the functional form of the hedonic price equation. First the distribution of income plays as strong a role in determining the functional form as preference parameters. Given any preference function, we can induce substantial changes in the form of the hedonic equation by changing the distribution of income. This result conforms with results of Rosen and Quigley and supports the use of best fit techniques. Further, one may take the hedonic equation as part of the household's exogenous budget constraint. Second, care must be taken in applying best fit techniques. While there is no necessity for the hedonic price equation to be convex, gross departures from convexity seem unlikely. It is possible for Box-Cox methods to yield many kinds of curvatures.

We have also learned some important lessons in the use of the hedonic price equation for welfare measurement. Despite excellent fits, hedonic price equations may not give good estimates of marginal bids. And Box-Cox estimation techniques do not necessarily yield hedonic price equations which have curvature appropriate for welfare measurement. This suggests a careful look at the distribution of marginal prices. How does the distribution of the marginal bid vary with parameter estimates from the hedonic price equation? This sort of question will be explored in detail in succeeding EPA work.

We have shown that for small changes in a single attribute, aggregated households' restricted willingness to pay is only a modest underestimate of the changes in rent. Further, we have shown that some attention must be paid to the distribution of income (and other household characteristics) in computing aggregate benefits.

[illegible]

of a simulation model in exploring the workings of hedonic markets. In additional work for EPA, we will use this approach with much more realistic data on housing markets to assess hedonic techniques.

CHAPTER 6

FOOTNOTES

1 McConnell is with the Department of Agricultural and Resource Economics of the University of Maryland. T. T. Phipps is with Resources for the Future, Washington, D. C.

2 In his empirical work, Quigley used a GCES utility function.

3 The coefficients for the model (6.6) were estimated via maximum likelihood using SHAZAM's 'BOX' routine.

APPENDIX, CHAPTER 6

Parameters of Simulation Model

Stone-Geary

j	0	1	2	3
β_j	.8	.06	.04	.1
δ_j	1000	5	15	20

Translog

$\beta_0 = 2$ $\beta_1 = .06$ $\beta_2 = .04$ $\beta_3 = .1$

δ_{ij}

i/j	1	2	3
1	-.3	.15	.2
2	.15	-.2	.5
3	.2	.5	-.1

The supplies of z were generated as follows:

z_1: uniform (6,26)

z_2: uniform (16,26)

z_3: uniform (21,31).

The distributions of income were generated as follows:

1. uniform [10,000, 30,000]
2. segmented uniform
   a. 125 observations uniform [5,000, 15,000]
   b. 125 observations uniform [20,000, 40,000]
3. Pareto generated as $y = y_0(1-u)^{\alpha}$ where $\alpha = {}^{-1}/_{1.2}$, $y_0 = 4000$, and u is uniform [0,1]
4. normal (20,000, $225 \cdot 10^6$).

Each distribution was transformed to have a mean of 20,000.

CHAPTER 7

SHOULD THE ROSEN MODEL BE USED TO VALUE ENVIRONMENTAL AMENITIES?

Maureen Cropper[1]

7.1 Introduction

There is a large literature in both urban and environmental economics which attempts to value site-specific amenities--access to workplace or air quality--using data on residential property values.[2] With few exceptions these studies appeal for their theoretical justification to Rosen's model of hedonic markets, and they follow his two-stage procedure in valuing amenities. In the first stage property values are regressed on housing characteristics and location-specific amenities to estimate an hedonic price function. The partial derivative of this function with respect to an amenity is interpreted as the marginal value which consumers place on the amenity. In the second stage marginal amenity price, computed from the hedonic price function, is regressed on the quantity of the amenities consumed and household characteristics to estimate a marginal willingness to pay function.

The purpose of this chapter is to discuss why these procedures may be inappropriate for valuing location-specific amenities, and why a discrete model of location choice may be preferred to the Rosen model on theoretical grounds. Reasons why the Rosen model may be inappropriate fall into three categories. First, some amenities are inherently discrete (whether a house has a river view), implying that the individual cannot make marginal adjustments in the amounts consumed. The assumption that marginal adjustments are possible, which is crucial to the Rosen model, is therefore unwarranted and renders the model inappropriate.

A second difficulty occurs when amenities which are in principle continuous assume only a few values in an urban area due to economies of scale in production. Examples of these include high school quality, which can [illegible] values as there are high schools, and quality of the local police force. The problem here is that local public goods, which require a minimum population for efficient production, cause indivisibilities in the set of amenities available (Ellickson, 1979). Thus, as with inherently discrete amenities, the individual cannot make marginal adjustments in quantities consumed.

These two problems, of course, are not unique to the attributes of locations. In markets for differentiated products, such as automobiles, one encounters inherently discrete attributes (the number of doors on a car) and finds "holes" in the menu of choices caused by economies of scale. (Only a few engine sizes are available to consumers due to the large amounts of product-specific capital required for engine manufacture.) The third difficulty with the Rosen model is, however, unique to the location choice problem.

A key assumption of Rosen's model is that each characteristic of a product can be varied independently of the others, subject only to a budget constraint. In the location choice problem, however, the attributes of location often cannot be varied independently of one another. This is because these attributes are tied to geographic location, and the choice of geographic location is a two-dimensional choice. Thus, if one wishes to model location choice as choice in amenities space, one must add the constraint that the choice of amenities $1,\ldots,g$ determines the amounts of amenities $g+1,\ldots,n$ consumed. Constraints of this type destroy the main result of the Rosen model, viz., that each amenity is consumed to the point where its marginal value to the consumer equals its marginal price. The two-stage procedure described above therefore cannot be applied.

The foregoing problems are discussed at length below. Section 7.2 reviews the Rosen model and discusses whether the model should be applied when some characteristics of goods are available only in discrete amounts. In Section 7.3 the model is applied to the choice of residential location. This means that geographic constraints must be added to the problem, and the section explores the implications of these constraints for location choice in amenities space. The difficulties discussed in Sections 7.2 and 7.3 can be resolved in part by estimating a discrete model of residential choice, in which the objects of choice are geographical locations. The structure of such models is outlined in Section 7.4. Section 7.5 concludes the chapter.

7.2 Consumer Choice in an Hedonic Market

In the model developed by Rosen to explain product differentiation under pure competition alternative brands of a product are indexed by an n-dimensional vector, $\underline{z}$, $\underline{z} \in R^n$, which describes the amount of each attribute provided by the brand. In the special case in which the consumer purchases only one unit of the brand his utility is a function of the vector $\underline{z}$ and the quantity consumed of a numeraire good, x,

$$U = U(x,\underline{z}). \tag{7.1}$$

U is assumed to be strictly increasing in x, strictly quasi-concave in $(x,\underline{z})$, and twice differentiable. Th[illegible] (7.1) subject to a budget constraint

$$p(\underline{z}) + x \leq y, \tag{7.2}$$

where y is income and $p(\underline{z})$, the hedonic price function, gives the unit cost of the differentiated commodity as a function of the attribute vector $\underline{z}$. In Rosen's presentation the set of $\underline{z}$'s available to the consumer is infinite and $p(\underline{z})$ is assumed to be differentiable.

For the present discussion two features of the model should be emphasized. One is that the consumer is free to choose each attribute of the brand independently of the others, subject only to his budget constraint. The other is that his choice set is infinite. Together these assumptions imply that the consumer equates the marginal value of each attribute to its marginal price,

$$\frac{\partial U/\partial z_i}{\partial U/\partial x} = \frac{\partial p}{\partial z_i}, \qquad i = 1,\ldots,n. \tag{7.3}$$

Equation (7.3) implies that the derivative of the hedonic price function with respect to amenity i equals the consumer's willingness to pay for that amenity at the level he is currently consuming. It also justifies the second stage of the Rosen procedure in which the coefficients of the marginal willingness to pay functions (the left-hand-sides of (7.3)) are estimated.[3]

In the notation of this section the problem of inherently discrete amenities occurs when some of the z_i's can assume only a countable number of values. For example, in choosing an oven the characteristic "fuel type" can assume only two values, gas or electric. Formally, suppose that z_1 can assume only two values but that the other z_i's are available in infinitely divisible quantities. In this case the marginal rate of substitution of z_1 for x is, of course, not defined, and (7.3) does not apply when i = 1. The choice of $\underline{z}$ is now a mixed discrete-continuous choice problem. Conditional on z_1, the remaining n - 1 equations in (7.3) can be solved together with (7.2) to yield conditional demand functions for x and for amenities 2,...,n. Upon substituting these functions in (7.1) one obtains an indirect utility function conditional on z_1, $V(z_1)$. The value of z_1 is selected which maximizes $V(z_1)$.

When z_1 is inherently discrete one is still interested in measuring the parameters of the utility function since willingness to pay for discrete changes in z_1 is well defined. This can be done by simultaneously estimating the last n - 1 equations of (7.3), the hedonic price function, and an equation for the probability of selecting z_1. Applying the Rosen model to discrete attributes, however, does not make much sense. The problem is not simply that the marginal willingness to pay function for z_1 is literally not defined, but that the notion of a marginal bid function assumes that the values of z_1 can be ordered. This is usually not the case with inherently discrete amenities, e.g., "fuel type" or "river view"; thus the Rosen model cannot be viewed as an approximation to reality in this case.

The case in which continuous attributes, such as school quality, happen to be available only in discrete amounts is somewhat different.[4] Although this problem is formally equivalent to the problem of inherently discrete attributes, and can be solved as a mixed continuous-discrete choice problem, it differs from the foregoing problem in one important respect: with attributes such as school quality the marginal willingness to pay function is a meaningful concept which one can try to approximate using the Rosen model.

To illustrate, suppose that z_1, the only amenity of interest, assumes three values within an urban area. The smooth curve pictured in Figure 7.1 is fitted to these three points, A, B and C, and the slope of the curve at each of

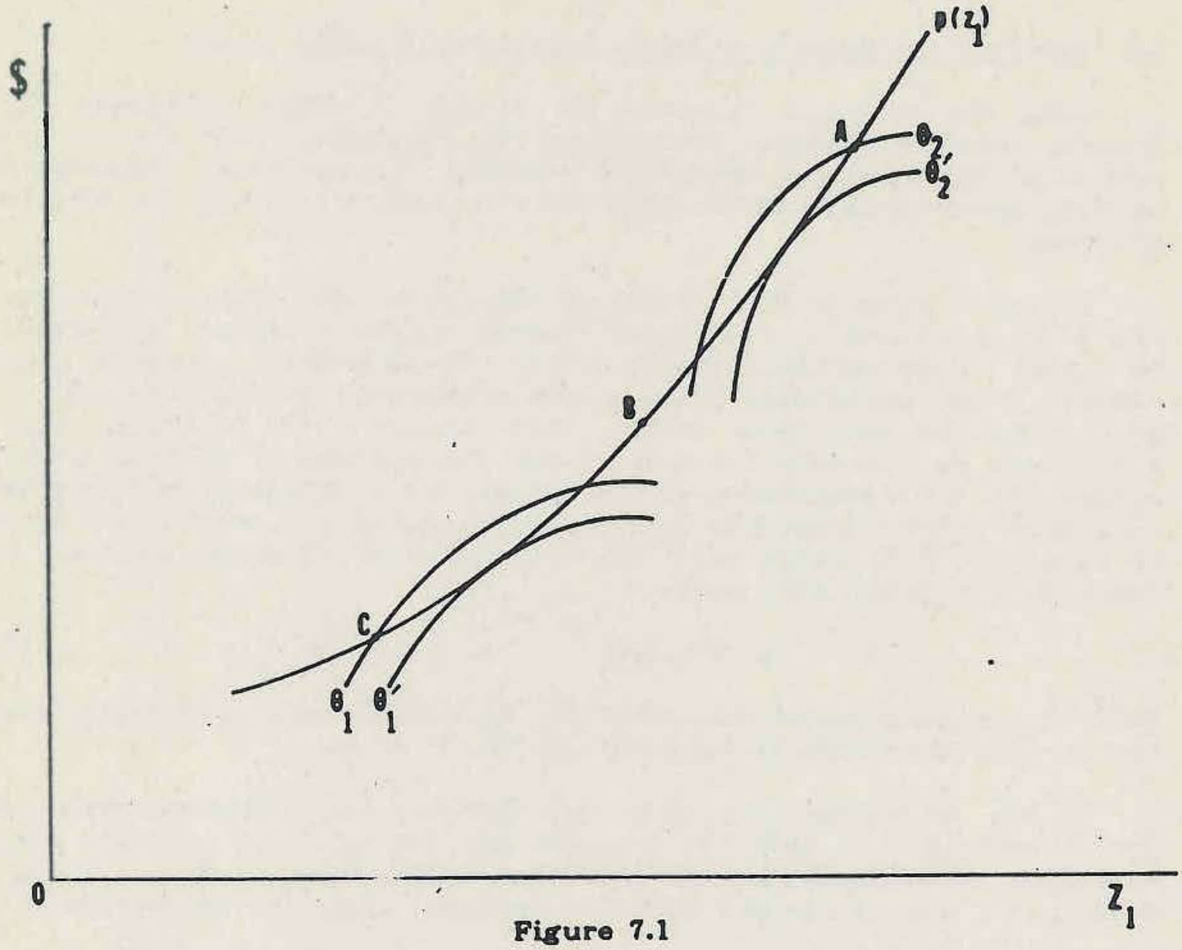


Figure 7.1

Bid Functions and Hedonic Price Functions

these points is interpreted as the marginal value of z_1 to the persons consuming that amount of the amenity. In reality, however, the slope evaluated at point C underestimates the marginal willingness to pay for z_1 by person 1, whose best choice of z_1 among the three alternatives is C. By contrast, the slope of the estimated hedonic price frontier overestimates marginal willingness to pay for person 2, whose optimal choice is point A. The failure of (7.3) to hold for persons 1 and 2 biases estimates of the marginal willingness to pay function; however, one suspects that this bias should diminish as the number of values of z_1 available increases. In this sense, one can justify the Rosen model as an approximation when there are "holes" in the data. This is not true when the amenity in question is inherently discrete.

7.3 Applying the Hedonic Model to Residential Location Choice

While the problems discussed in Section 2 create difficulties in using Rosen's model to measure preferences for attributes, they are not problems unique to the choice of residential location. The problems discussed in this section, however, have few counterparts in hedonic markets for manufactured products.

The main point of this section is that when the model of equations (7.1) and (7.2) is applied to residential location choice additional constraints must be placed on the problem because of the two-dimensional nature of geographic choice. These constraints prevent the household from independently varying all n amenities and thus render (7.3) invalid. To emphasize that these constraints do not arise because of the discreteness of available choices, we assume that all n location-specific amenities are available in infinitely divisible quantities.[5] Even when this is true, the choice of $\underline{z}$ is constrained by the set of equations (7.4) which describes the vector of amenities available at each point (u,v) in geographic space,

$$z_i = f_i(u,v) \qquad i = 1,\ldots,n. \tag{7.4}$$

Since the amenity vector consumed can be altered only by changing locations, the set of available $\underline{z}$'s is implicitly defined by (7.4).

To see intuitively why (7.4) may prevent the individual from independently varying all n amenities suppose that two of the n amenities are access amenities. Specifically, let z_i = distance to the point (u_i,v_i), $i = 1,2$, where (u_1,v_1) and (u_2,v_2) are two points of interest, (e.g., the workplaces of a two-

and since the circumferences of two distinct circles intersect in at most two points, there are at most two points in the u-v plane corresponding to any feasible (z_1,z_2) pair (see Figure 7.2).[6] This implies that once z_1 and z_2 are determined the individual has at most two choices for each of the remaining n-3 amenities of interest.[7]

The necessary conditions for location choice in amenity space are therefore not given by (7.3) if two or more amenities are access amenities. For z_1 and z_2 defined as above the household would locate two points in geographic (and amenity) space by choosing z_1 and z_2 to maximize (7.1)

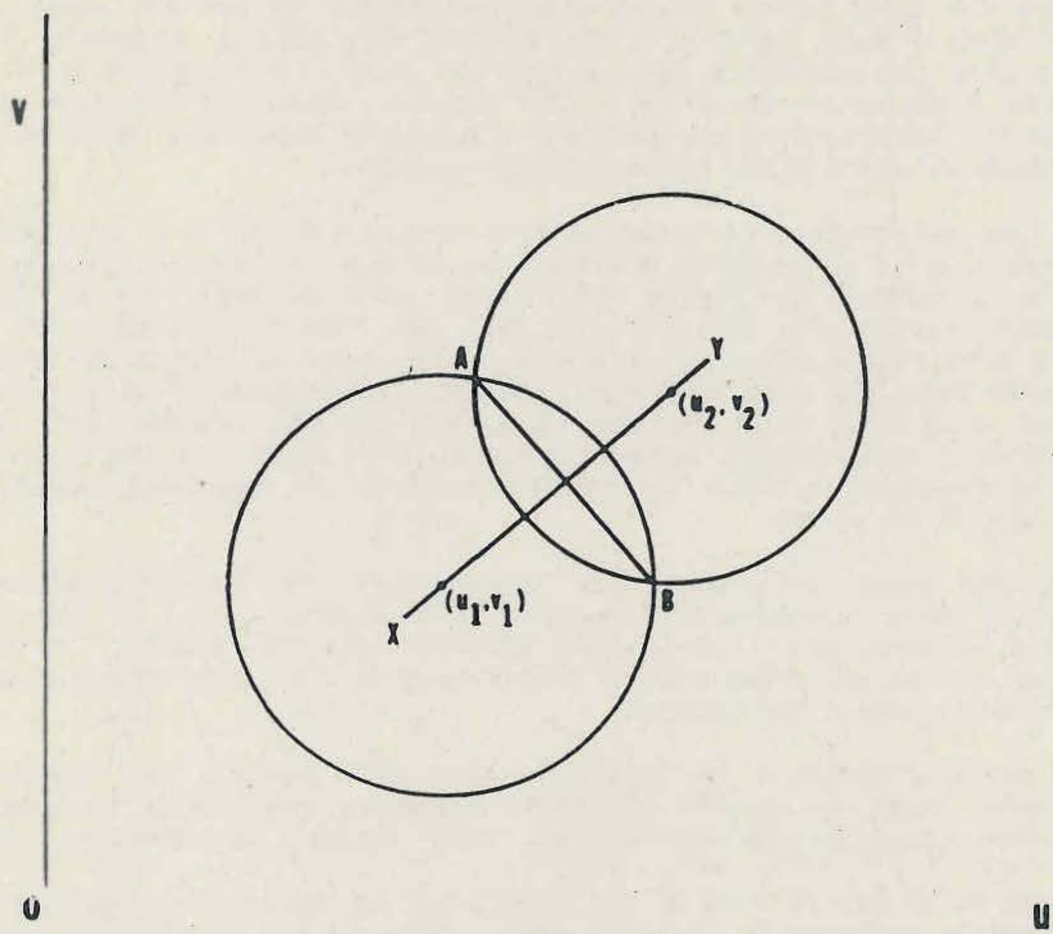


Figure 7.2

Locational Restrictions on Choice

subject to (7.2), (7.4) and a feasibility constraint, (7.5),

$$A_1 \cap A_2 \neq 0 \tag{7.5}$$

where $A = \{(u,v) | z_i^2 = (u - u_i)^2 + (v - v_i)^2\}$, $z_i = 1,2$. The household would then locate at the point yielding the higher utility.

This example gives a specific and reasonable instance of the way in which the two-dimensional nature of location choice limits choice in amenities space. Suppose, however, that access amenities are not of interest to a household. Is choice in amenities space still restricted by the two-dimensional nature of geographic choice? The answer to this question depends on the nature of the functions $f_i(u,v)$, i.e., on the distribution of amenities over geographic space. Consider the level curves of two amenities plotted in the u-v plane. If the distribution of each amenity is monocentric and radially symmetric then its level curves are concentric circles and the same result obtains as when the amenities are access amenities: any feasible choice of the two amenities restrict the household to two points in geographic space and hence to at most two values for each of the remaining n-2 amenities.

If the distribution of an amenity is asymmetric or if it is multicentric, then the number of possible intersections of any two level curves increases. This is illustrated in Figure 7.3, which pictures level curves for total suspended particulates and distance from the CBD in Baltimore, MD. It is evident from Figure 7.3 that the choice of 60 $\mu g/m^3$ of particulate matter and five miles from the CBD no longer restricts the household to two locations; however, only four points satisfy these two amenity values. For amenities that occur in continuously variable amounts, it is clear that the choice of two or three amenities restricts the choices available for remaining amenities to a finite number of points.

At this point the reader may wonder how the foregoing argument is altered if some site-specific amenities are discrete, e.g., if the relevant pollution variable is an index which assumes only five values. In this case the level curves are areas and no longer restrict the choice of other amenities in the manner described above.[8]

It should, however, be borne in mind that for the two dimensions of geographic space to restrict choice in amenities space it is necessary that only two amenities be continuous, with spatial distributions that are approximately symmetric and monocentric. This [illegible] condition to satisfy in view of the importance of "distance to work" in household location decisions.[9] In a two-earner household it is certainly reasonable that distance to each person's place of work is an important amenity in so far as residential location is concerned.[10]

7.4 Discrete Models of Residential Location Choice

Although conceptually different, each of the three problems described above has a similar effect on the household's choice of amenities: it causes the choice set to become discrete (at least for some subset of amenities) thus

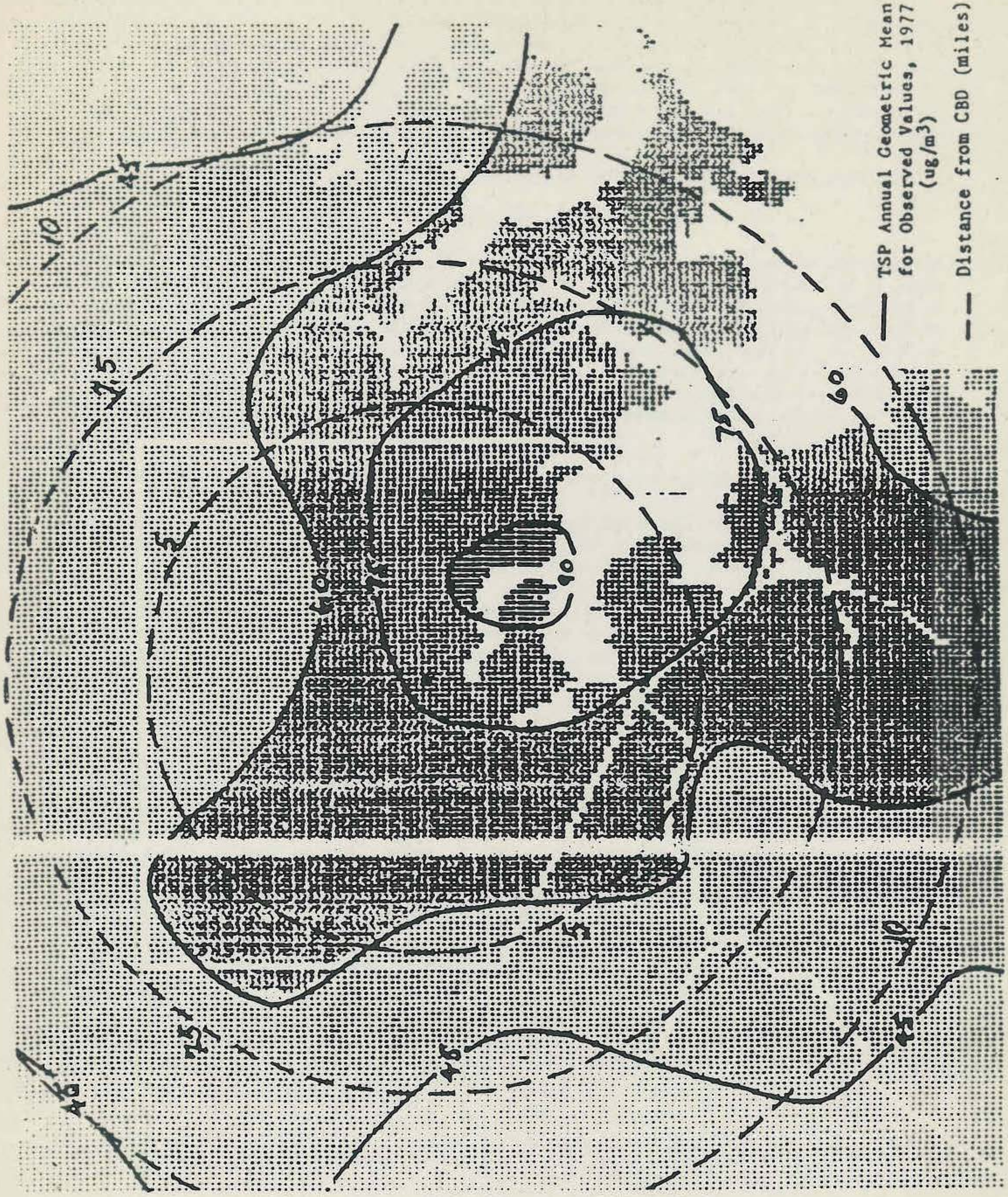

TSP Annual Geometric Mean for Observed Values, 1977 (ug/m3)
Distance from CBD (miles)

Figure 7.3

violating the assumptions of the Rosen model.

This suggests that one consider discrete choice models of residential location as a method of valuing site-specific amenities. In a discrete model of residential location the objects of choice are geographic locations, indexed i, where the set of all i is finite. To each location there corresponds a vector of amenities $\underline{z}_i$. As in the Rosen model, utility is defined over $\underline{z}$ and a numeraire x. By making locations rather than amenities the objects of all choice, geographic restrictions are incorporated into the problem <u>ipso facto</u>.

In this framework, household h chooses the location i for which

$$U_{ih} = U_h(\underline{z}_{ih}, y_h - p_i)$$

is highest, where p_i is the price of location i. To make the model a statistical one, it is usually assumed that utility is random from the viewpoint of the researcher since he cannot observe all attributes of locations. Redefining $\underline{z}_{ih}$ to include only those attributes observable by the researcher, utility may be written as the sum of a deterministic term, $V_{ih}(\underline{z}_{ih}, y_h - p_i)$, and a random term ε_{ih}. V_{ih}, also termed "strict utility," is usually written as a linear-in-parameters function of $y_h - p_i$, $\underline{z}_{ih}$ and interacts between these variables and household characteristics. The probability that household h selects location i is given by

$$P(V_{ih} + \varepsilon_{ih} > V_{jh} + \varepsilon_{jh}, \text{ all } j \neq i). \tag{7.6}$$

To value site-specific amenities given data on residential location choices one maximizes a likelihood function with individual terms of the form (7.6). If the $\{\varepsilon_{ih}\}$ are assumed to be identically distributed for all i and h with a Type I Extreme Value distribution, the resulting likelihood function corresponds to the multinomial logit model. If choice of house is also observed, a nested multinomial logit model is usually assumed (McFadden, 1978). Given estimates of the parameters of V_{ih}, random counterparts of compensating and equivalent variations can be constructed for changes in the $\underline{z}$ vector (see Hanemann (1984)).

7.5 <u>Conclusion</u>

The purpose of this chapter has been to explain why the Rosen model may be inappropriate for valuing location-specific amenities, such as air quality and local [illegible] If [illegible] amenities are inherently discrete (e.g., z_1 = location has a view of the beach), and if these discrete variables cannot be ordered, then the notion of a continuous bid function for amenities is meaningless, even as an approximation. In this case the Rosen model is clearly inappropriate. A second but less damaging situation occurs when amenities which enter the utility function as continuous variables are available only in discrete quantities for one reason or another. In this case one can at least view the Rosen model as an approximation to reality, which improves as the size of the discrete choice set increases.

The third situation emphasized in this chapter occurs when the two-dimensional nature of location choice restricts choice in amenities space. Here the amenities of interest enter the utility function as continuous variables and are also available in infinitely divisible quantities; however, the choice of two or more amenities restricts the number of choices available for the remaining amenities to a few. Since bid functions for location-specific amenities are defined in case three, it is tempting to use the Rosen model as an approximation to reality, as one might do in case two. This, however, is not possible. In case two, equation (7.3) at least may be viewed as holding approximately (see Figure 7.1). In case three, however, the first-order conditions of the Rosen model no longer apply since all n amenities cannot be chosen independently of one another.

The three situations described above argue for the use of a discrete choice model to value location-specific amenities. In the first and second situations the case for a discrete choice model is obvious. In the third it has been demonstrated that the choice of certain site-specific amenities restricts the household to a few points in geographic space and, hence, to a finite number of amenity vectors. The reader, however, may object that a discrete choice model is awkward when the number of choices is large, and that a commonly used discrete choice model, the multinomial logit, is flawed by the assumption that the error terms are independently and identically distributed.[11] There are several responses to these criticisms.

The fact that the number of possible residential locations is large may be considered a problem for two reasons, one computational and the other behavioral. The computational problem has been treated by McFadden (1978) who demonstrates that for purposes of estimating the multinomial logit model each household's choice set can be obtained by sampling from the universal choice set. Thus the existence of thousands of choices in the universal choice set need not pose a barrier to estimation.

The more disturbing problem created by a large choice set is behavioral. When the choice set is large it is unrealistic to assume that the individual compares all possible alternatives according to each attribute of interest. This limitation of discrete choice models can be overcome in two ways. If the choice set has tree structure (e.g., the household selects an area of the city, then a neighborhood, then a house), one can apply Tversky's hierarchical elimination-by-aspects model (Maddala). In this model the individual selects a single branch at each level of the decision tree, thus eliminating all alternatives not included in that branch. [illegible] by Cha, is to assume that the individual ranks alternatives according to a small subset of attributes and then compares only the k highest ranked alternatives according to all attributes.

The assumption that the random component of utility is independently and identically distributed across households and alternatives is most objectionable when the objects of choice are individual houses rather than large neighborhoods. For example, it is unlikely that the unobserved attributes of a house are distributed independently of those of the house next to it. Correlation between the unobserved attributes of alternatives on the lower levels of a decision tree is, however, allowed in McFadden's (1978) nested logit model. Thus, the Independence of Irrelevant Alternatives property need not destroy discrete choice models.

One final point. Although it would be foolish to pretend that discrete choice models are not without econometric difficulties, these difficulties must be judged in light of the econometric problem of the Rosen model, described in earlier chapters. From this perspective discrete choice models are a method of valuing environmental amenities worthy of consideration.

CHAPTER 7

FOOTNOTES

1 Department of Economics, University of Maryland.

2 Portions of this literature have been summarized by Freeman (1979a), Diamond and Tolley, and Bartik and Smith.

3 For the coefficients of the marginal willingness to pay functions to be estimated efficiently, these functions must be estimated jointly with the hedonic price function.

4 In the introduction the fact that some amenities are available only in discrete amounts was motivated by economies of scale in the provision of local public goods. An analogous problem occurs if attributes which are available in infinitely divisible amounts are coded as discrete by data collectors.

5 The consequences of relaxing this assumption are explored below.

6 There must be at least one point in the u-v plane corresponding to (z_1,z_2) or the (z_1,z_2) pair is not feasible.

7 If there is a third point of interest in the city, (z_3, defined analogously to z_1 and z_2) the above argument is even stronger. As long as the three points of interest in the city do not lie on the same straight line it can be shown (see Appendix) that any feasible choice of (z_1,z_2,z_3) uniquely determines the household's geographic location. Once location is determined the levels of all other amenities are uniquely given by (7.4) since there is only one value of z_i at each point in geographic space.

8 In Figure 7.2, for example, the area between 45 and 60 $\mu g/m^3$ might represent a single value of the pollution index.

9 Empirical studies of residential location choice (Anas, 1982; Lerman, 1979) have consistently found distance (or travel time) to work to be a statistically significant determinant of household location. One difficulty in assessing the importance of distance to work within the Rosen framework is that any amenity which varies with household as well as location cannot be valued unless all households are similar. Thus, in an urban area with many work centers, distance to work center i may not have a statistically significant coefficient in an hedonic price function area, even though distance to work is an important determinant of residential location.

10 This assumes, of course, that workplace location is fixed as far as the residential location decision is concerned. If workplace location is determined jointly with residential location then the argument of Figure 7.2 must be applied to each workplace location. As long as the number of possible workplace locations is finite the choice of z_1, z_2, (u_1,v_1) and (u_2,v_2) still restricts the choice of amenities $z_3,\ldots,z_n$ to a finite number of points.

11 This assumption together with the assumption that each error term has a type I Extreme Value distribution gives rise to the Independence of Irrelevant Alternatives property of the multinomial logit model. This means that the probability of selecting alternative i divided by the probability of selecting alternative j is independent of the other alternatives available.

APPENDIX TO CHAPTER 7

The purpose of this appendix is to prove that any feasible choice of amenities z_1, z_2 and z_3 where z_i = distance to the point (u_i,v_i), i = 1,2,3, uniquely determines a household's location in the u-v plane, provided that all of the points (u_i,v_i), i = 1,2,3, do not lie on the same straight line. For any z_i the locus of points z_i away from (u_i,v_i) form the circumference of a circle with radius z_i. The result to be proved is that the circumferences of the three circles which are z_i away from (u_i,v_i), i = 1,2,3, intersect in at most one point, provided the points (u_i,v_i), i = 1,2,3, do not lie on the same straight line. If the three circumferences do not intersect in at least one point then the choice of (z_1,z_2,z_3) is not feasible.

We begin by noting that the circumferences of any two distinct circles intersect in at most two points. Call these points A and B and let AB denote the line joining A and B. (See Figure 7.2.) The line joining the centers of the two circles must be perpendicular to AB. If A and B lie on the circumferences of two circles then the center of each circle must be equidistant from A and B. The locus of points equidistant from any two points is a line perpendicular to the line joining the two points. Call this line XY.

For a third circle to intersect the first two in more than one point it must pass through points A and B. We show that this can happen if and only if the center of this circle lies on the line XY. To see that this is possible only if the center of the third circle lies on XY note that the circle whose circumference passes through points A and B must, by definition, be equidistant from A and B. However, the locus to points equidistant from any two points is a line perpendicular to the line joining the two points. Thus, the circumference of three circles can intersect in more than one point only if their centers lie on the same straight line.

CHAPTER 8

SUMMARY AND ASSESSMENT

8.1 Introduction

The purpose of the hedonic component of the Maryland-EPA Cooperative Agreement, as originally defined, was to "solve the identification problem in hedonic models." Our conclusions concerning the identification problem, based on the reasoning of Chapters 3 and 4, in hedonic markets can be solved only be assuming fairly specific functional forms for preferences and the hedonic price equation, without the ability to test whether these forms hold. While there may be occasions when household behavior conforms with the necessary assumptions, the difficulties in statistically testing such assumptions make the solution to the identification problem rather unsatisfactory. Because we have concluded that identification of preference parameters is quite difficult, we have also explored other issues in hedonic models and other methods of assessing the benefits of environmental improvement from housing transactions.

8.2 The Identification Problem: Summary and Resolution

Two questions arise in addressing the issue of the identification problem. The first pertains to whether a solution exists. The second relates to the costs of the solution.

8.2A. Can We Do It?

The identification problem deals with the question: can we use the hedonic model to recover information about preferences? In particular, can the parameters of the preference function be identified and therefore used for determining the benefits of non-marginal changes in attributes? The answer to the basic question of identification is 'yes', we can identify the parameters of preference functions under certain conditions. For participants in a single market who face the same hedonic price equation, we can identify their preference parameters in the following way:

(1) Impose enough structure on the linear in parameters model so that it can be shown to be identified by traditional exclusion criteria (Section 4.3A). The variables excluded will typically be nonlinear transformations of endogenous variables.

(2) Successfully estimate the whole system of equilibrium conditions using maximum likelihood methods (Section 4.3B). Successful estimation implies that preferences and the hedonic price equation have sufficiently different curvature to allow the maximum likelihood estimates to converge.

(3) Estimate the reduced form with attributes as endogenous variables and show that the preference parameters can be derived uniquely

from the reduced form parameters (Section 4.3C, Appendix 4.A). This can be achieved in a very limited number of cases.

(4) Estimate different linear hedonic price equations from segmented markets or multiple markets and use the coefficients as prices in a traditional demand system with prices as parameters (Section 4.4B).

Finally, for households in different markets, we have an additional approach:

(5) Use marginal prices from multiple-cities hedonic price equations, and estimate the system as Rosen originally intended.

Of the five suggested approaches, only the last makes use of the traditional Rosen two step model. Further while multiple markets may provide the basis for identification, the numerical questions of how many markets one needs and what additional structure must be imposed remain to be investigated. It is worth emphasizing that regardless of the chosen functional form, there is no way to determine identification from the simple application of the Rosen two step approach in the single market setting. This holds even when we derive the marginal value functions from an explicit utility function as, for example, in Quigley (1982). (See Appendix 4.A, example 2.)

Identification of parameters in an equation is always derived from prior information. In some cases the imposition of prior information is innocuous in that it has no behavioral implications. For example, the normalization of the parameter on the dependent variable in a single equation linear regression model is necessary for the estimation of the model but has no behavioral implications. On other occasions the imposition of prior information has behavioral implications, but is quite plausible. For example the structural parameters of a model of an agricultural commodity might be identified by the plausible assumptions that demand is increased by increases in per capita income and supply is increased by greater summer rainfall.

The resolution of the identification problem in hedonic models is less satisfactory. In all of the five approaches to identification given above, there are no simple and intuitive assumptions, such as rainfall influences supply but not demand, to identify parameters. No such assumptions are available because the basic equations which are simultaneous stem from the same actors--the individual households from which the data are taken. Instead, the identification of preference parameters in hedonic models comes only as a result of assumptions about functional form. We have shown, for example (in equations 4.15 ... 4.21) that the specification of the hed[illegible] cubic rather than a quadratic function will serve to identify a linear marginal rate of substitution function. While in some cases such assumptions about functional forms are subject to nested testing (for example when the hedonic price equation is recursive), in most cases they are not. Most important, such assumptions have none of the compelling plausibility that identifies the demand for an agricultural commodity by omitting summer rainfall. In sum, we can identify the parameters of preferences, but only by imposing assumptions about functional form for preferences and for the hedonic price equation which rarely have any intuitive appeal.

In one sense, this result does not make identification in hedonic models quite as gloomy a prospect as it seems. Functional forms are not devoid of economic content. The general requirement for household equilibrium in the Rosen model is for the preference function to show more concavity than the hedonic price equation. The second order conditions have a certain economic force. However, such economic content typically requires functions nonlinear in parameters, and thus ignores some fairly significant practical hurdles. Identification of models nonlinear in parameters requires successful estimation by maximum likelihood, an unrealistic requirement for the typical model with many attributes. And converting to linear-in-parameter models by polynomial approximation usually obscures the economic content of functional form. Thus, practical reasons undermine the economic content of functions.

Thus we are in a position to identify the preference parameters of hedonic models, by imposing structure on the marginal rate of substitution functions and on the hedonic price equation. Typically the assumptions needed to induce identification will be fairly severe and arbitrary, but if identification gives us enough new information such assumptions may well be worthwhile.

In sum, identification of parameters of preference functions in hedonic models can be achieved through assumptions about functional form. Such assumptions are commonly made in empirical work, but they are generally testable. In the hedonic model, they are typically not testable. Further, the gains in accuracy do not seem to be worth it. If we use the hedonic model for welfare changes, we may as well use the guidelines for approximations laid by Freeman ten years ago (Freeman, 1974a).

8.2B. <u>Is It Worth It</u>?

Whether identification, when conceptually feasible, is worthwhile depends in part on whether the implied behavior is plausible. Thus an important question in the context of identification is not whether the appropriate coefficients can be recovered, but whether the prior restrictions imply plausible behavior. As noted above, standard commodity models are identified typically by appealing to constraints on behavior: the level of rainfall does not affect the demand for wheat. What sort of behavior is implied by the methods of identification implied by this volume?

First consider the hedonic price equation. The results of both Chapter 3 and [illegible] hedonic price equation can help identify the marginal rate of substitution equations. But other theory (Rosen, Quigley) as well as the empirical results of Chapter 6 demonstrate that no particular behavior can be deduced from curvature of the hedonic price equation. As we showed in detail in Chapter 6, the preference parameters, the distribution of household tastes, and the distribution of amenities determine jointly the functional form of the hedonic price equation. Further, we typically have no strong prior beliefs about this functional form, but are free to estimate best fitting functional forms. Thus, part of the solution to the identification problem comes from the functional form of the hedonic price equation, and only in rare instances can we ascertain the behavioral implications of such forms.

Results from Chapter 4 suggest that identification is likely to be enhanced by separability assumptions. That is, identification of one marginal rate of substitution function is easier when it excludes variables which appear in other marginal rate of substitution functions. Such exclusion of variables occurs when the utility function is separable. This result is in keeping with the literature on the estimation of demand systems, where it has long been recognized that various forms of separability would reduce the estimation burden. There is a crucial distinction, however, between assuming separability to reduce the number of parameters to be estimated in demand systems and assuming separability to *identify* parameters in hedonic markets. In demand systems, we can test for separability. In hedonic markets, we cannot typically test for the assumption of separability, for without it, we do not even have preference parameters.

In the end, identification of preference parameters in hedonic markets requires assumptions of unknown validity in the hedonic price equation and separability in the preference function. Our knowledge of behavior is not sufficient for us to argue that separability of the preference function is plausible.

8.3 Suitability of the Rosen Model for Valuing Environmental Amenities

Chapter 7 has questioned, whether the Rosen model can be used for environmental quality. In Rosen's model, which was developed to explain product differentiation, brands are indexed by an n-dimensional vector of attributes. In selecting a brand the consumer is faced with an infinite set of attribute vectors and can choose each attribute of the brand independently of the others, subject only to his budget constraint. Utility maximization thus requires that the marginal utility of each attribute ($\partial U/\partial z_i/\partial U/\partial x$) be equated to its marginal price. This justifies the interpretation of the partial derivative of the hedonic price frontier as measuring the marginal value of an attribute to some consumer.

Consumers in the land market, however, do not have as many degrees of freedom as purchasers of manufactured products. Even when the set of residential locations is infinite, so that marginal changes in location can be made, the consumer cannot freely vary each of n attributes of the housing site. This is because the consumer has only two degrees of freedom in choosing a residential site -- he can select its latitude and longitude. In making marginal changes in latitude and longitude the consumer must weigh the effect of these changes on each of the n attributes of the housing site and compare a weighted sum of marginal valuations to the mnrginal valuations to the marginal cost of the move. The consumer is therefore unable to equate the marginal value of each attribute to its price, and the slope of the hedonic price frontier with respect to an attribute cannot be interpreted as the marginal value of the amenity to the consumer. Since the consumer cannot freely choose all elements of the attribute vector his demand (bid) functions for various attributes will not correspond to those in Rosen's model.

Since much of the empirical work which attempts to value air quality follows Rosen's approach, these studies must be re-evaluated. One way of doing this is to be compare the results of these studies with the results of alternative approaches suggested in Chapter 7.

8.4 Future Research

Our research on the hedonic model has focused on two issues: the identification problem and the use of the Rosen model for environmental amenities. In the second cycle of our Cooperative Agreement, we plan to explore these issues in several different ways. First, we plan to pursue approaches which emphasize discrete choices or bids for housing. The bidding approach will follow the work of Ellickson (1981), Lerman and Kern (1983), and Horowitz (1983). The discrete choice models will follow the work of McFadden (1978) and Anas (1982). We plan to develop and estimate a variety of these models on several different data sets, with the emphasis on measuring the benefits of improvements in air quality. In the process of developing new approaches, it will be useful to compare these empirical results with empirical results from the Rosen model.

While the departure from the Rosen model means a loss of some intuitively appealing properties such as continuity and equilibrium at the margin, it also gives us the opportunity to discard or at least test two maintained but unrealistic hypothesis: perfect information and equilibrium. The discrete choice or bid models of Horowitz, McFadden and others do not require an equilibrium in the housing market. Further they do not require complete information. Hence this research direction not only allows us to advance from a model which does not seem to fit the residential housing market in concept. It also allows us to model the actual purchase or rental of a housing unit in a much more plausible way.

Second, we plan to use the simulation approach of Chapter 6 to explore more workings of the hedonic model. We will enrich the simulation approach so that we are modelling the housing markets of discernible cities. In particular, we will attempt to mimic the behavior of markets in Los Angles and Baltimore. Further, we will develop markets in a fair number of cities to see if we can determine in what circumstances the multiple markets approach to identification will work.

Finally, we have concluded that because there is a feasible but perhaps not worthwhile solution to the identification problem it is [illegible] to proceed with benefit estimation cautiously using the slopes of the hedonic price equation. We will explore the statistical characteristics of these slopes for different forms of hedonic price equations.

Adelman, I. and Z. Griliches. 1961. "On an Index of Quality Change," Journal of the American Statistical Association 56:535-548.

Alonso, W. 1964. Location and Land Use. Cambridge: Harvard University Press.

Anas, A. 1982. Residential Location Markets and Urban Transportation: Economic Theory, Econometrics and Policy Analysis with Discrete Choice Models. New York: Academic Press.

Anderson, R. J. and T. Crocker. 1971. "Air Pollution and Residential Property Values." Urban Studies 8:171-180.

Barnett, W. A. 1977. "Pollak and Wachter on the Household Production Function Approach." Journal of Political Economy 85:73-82.

Bartik, T. J. 1983. "The Estimation of Demand Parameters in Hedonic Price Models." Nashville: Vanderbilt University Dept. of Economics.

Bartik, T. J. and V. K. Smith. 1984. "Urban Amenities and Public Policy." Working paper. Department of Economics and Business, Vanderbilt University.

Bayless, M. 1983. "Measuring the Benefits of Air Quality Improvements: A Hedonic Salary Approach." Journal of Environmental Economics and Management 9:81-99.

Becker, G. 1965. "A Theory of the Allocation of Time." Economic Journal 75:493-517.

Bender, B., T. J. Gronberg and H. Hwang. 1980. "Choice of Functional Form and the Demand for Air Quality." Review of Economics and Statistics 62:638-643.

Blackburn, A. J. 1971. "Equilibrium in the Market for Land: Obtaining Spatial Distributions by Change of Variable." Econometrica 39:641-644.

Blomquist, G. and H. Worley. 1981. "Hedonic Prices, Demands for Urban Housing Amenities, and Benefit Estimation." Journal of Urban Economics 9:212-221.

Bockstael, N. E. and K. E. McConnell. 1983. "Welfare Measurement in the Household Production Framework." American Economic Review 73:806-814.

Bowden, R. 1973. "The Theory of Parametric Identification." Econometrica 41:1069-1074.

Bradford, D. and G. Hildebrandt. 1977. "Observable Preferences for Public Goods." Journal of Public Economics 8:111-131.

Brookshire, D. S., M. Thayer, W. Schultz and R. C. d'Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches." [illegible]

Brookshire, D. S., M. A. Thayer, J. Tschirhart and W. D. Schulze. 1985. "A Test of the Expected Utility Model: Evidence from Earthquake Risks." Journal of Political Economy 93:369-389.

Brown, G. M., Jr. and R. Mendelsohn. 1984. "The Hedonic-Travel Cost Method." Review of Economics and Statistics 66:427-433.

Brown, G. M., Jr. and H. D. Pollakowski. 1977. "Economic Valuation of Shoreline." Review of Economics and Statistics 59:272-278.

Brown, J. N. 1983. "Structural Estimation in Implicit Markets" in (J. Triplett ed.) The Measurement of Labor Cost. NBER Studies in Income and Wealth, No. 48. Chicago: University of Chicago Press. 123-151.

Brown, J. N. and H. S. Rosen. 1982. "On the Estimation of Structural Hedonic Price Models." Econometrica 50:765-768.

Cha, D-D. 1983. "Disaggregate, Behavioral, Two-State Travel Demand Model: Choice Set Identification and Choice Process." Doctoral Dissertation, University of Maryland.

Court, L. M. 1941. "Entrepreneurial and Consumer Demand Theories for Commodity Spectra." Econometrica 9:135-162, 241-297.

Desvousges, W. H., V. K. Smith and M. P. McGivney. 1983. A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements prepared for U.S. EPA by Research Triangle Institute.

Diamond, D. B. and B. Smith. 1985. "Simultaneity in the Market for Housing Characteristics." Journal of Urban Economics 17:270-292.

Diamond, D. B. and G. S. Tolley (ed.). 1982. The Economics of Urban Amenities, Academic Press, N.Y.

Ellickson, B. 1979. "Local Public Goods and the Market for Neighborhoods" in (D. Segal ed.) The Economics of Neighborhood. New York: Academic Press.

Ellickson, B. 1981. "An Alternative Test of the Hedonic Theory of Housing Markets." Journal of Urban Economics 9:56-79.

Epp, D. J. and K. S. Al-Ani. 1979. "The Effects of Water Quality on Rural Residential Property Values." American Journal of Agricultural Economics 61:529-533.

Epple, D. 1982. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." Pittsburg: Carnegie-Mellon University.

Fisher, F. M. 1966. The Identification Problem in Econometrics. New York: McGraw-Hill.

Fisher, F. M. and K. Shell. 1968. "Taste and Quality Change in the Pure Theory of the True Cost-of-Living Index" in (J. N. Wolfe ed) Value, Capital, and Growth. Chicago: Aldine Publishing Co.

Freeman, A. M., III. 1971. "Air Pollution and Property Values: A Methodological Comment." Review of Economics and Statistics 53:415-416.

Freeman, A. M., III. 1974a. "On Estimating Air Pollution Control Benefits from Land Value Studies." Journal of Environmental Economics and Management 1:74-83.

Freeman, A. M. III. 1974b. "Spatial Equilibrium, the Theory of Rents and the Measurement of Benefits from Public [illegible] Journal of Economics 88:470-473.

Freeman, A. M., III. 1979a. The Benefits of Environmental Improvement: Theory and Practice. Baltimore: Johns Hopkins Press.

Freeman, A. M., III. 1979b. "Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues." Scandanavian Journal of Economics 81:154-173.

Freeman, A. M., III. 1982. Air and Water Pollution Control: A Benefit-Cost Assessment. New York: Wiley-Interscience.

Freeman, A. M., III. 1984. "The Hedonic Price Technique and the Value of Climate as a Resource" in (V. K. Smith and A. D. Witte, eds.) Advances in Applied Micro-Economics V. 3. Greenwich, Conn.: JAI Press, 169-186.

General Accounting Office. 1984. "Cost-Benefit Analysis Can Be Useful in Assessing Environmental Regulations, Despite Limitations." Report to the Congress.

Goldfeld, S. M. and R. E. Quandt. 1972. Nonlinear Methods in Econometrics Amsterdam-London: North Holland.

Griliches, Z. 1961. "Hedonic Price Indexes for Automobiles: An Econometric Analysis" in The Price Statistics of the Federal Government National Bureau of Economic Research: New York. Reprinted in Griliches (1971) pp. 55-87.

Griliches, Z. (ed.) 1971. Price Indexes and Quality Change Cambridge, Mass: Harvard University Press.

Hall, R. E. 1973. "Wages, Income, and Hours of Work in the U. S. Labor Force," in (G. Cain and H. Watts ed.) Income Maintenance and Labor Supply. Chicago: Markham.

Halvorsen, R. and H. O. Pollakowski. 1981. "Choice of Functional Form for Hedonic Price Equations." Journal of Urban Economics 10:37-49.

Hanemann, W. M. 1981. "Quality and Demand Analysis." manuscript, Department of Agricultural and Resource Economics, University of California, Berkeley.

Hanemann, W. M. 1984. "Applied Welfare Analysis with Qualitative Response Models," working paper, Department of Agricultural and Resource Economics, University of California, Berkeley.

Harrison, D. and D. L. Rubinfeld. 1978. "Hedonic Housing Prices and the Demand for Clean Air." Journal of Environmental Economics and Management 5:81-102.

Horowitz, J. L. 1983. "A Bidding Model of Urban Housing Markets with Disequilibrium and Incomplete Information." Prepared for 30th North Amesican meeting of the Regional Science Association.

Houthakker, H. S. 1952. "Compensated Changes in Quantities and Qualities Consumed." Review of Economic Studies 19:155-164.

Irish, M. 1980. "Optimal Quality Choice: An Empirical Study." Oxford Bulletin of Economics and Statistics 42:65-80.

Jud, G. D. and J. M. Watts. 1981. "Schools and Housing Values." Land Economics 57:459-470.

King, A. T. 1976. "The Demand for Housing: A Lancastrian Approach." Southern Economic Journal 40:1077-1087.

Lancaster, K. 1966. "A New Approach to Consumer Theory." Journal of [illegible]

Lerman, S. R. 1979. "Neighborhood Choice and Transportation Services" in (D. Segal ed.) The Economics of Neighborhood. New York: Academic Press.

Lerman, S. R. and C. R. Kern. 1983. "Hedonic Theory, Bid Rents and Willingness-to-Pay: Some Extensions of Ellickson's Results." Journal of Urban Economics 13:358-363.

Li, M. and H. J. Brown. 1982. "Micro-neighborhood Externalities and Hedonic Housing Prices." Land Economics 57:125-141.

Lind, R. C. 1973. "Spatial Equilibrium, the Theory of Rents, and the Measurement of Benefits from Public Programs." Quarterly Journal of Economics 87:188-207.

Linneman, P. 1981. "The Demand for Residence Site Characteristics." Journal of Urban Economics 9:129-148.

Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Economics. Cambridge, Cambridge University Press.

Maler, K. G. 1971. "A Method of Estimting Social Benefits From Pollution Control." Swedish Journal of Economics 88:111-131.

Maler, K. G. 1974. Environmental Economics - A Theoretical Inquiry Johns Hopkins Press: Baltimore.

McFadden, D. 1978. "Modelling the Choice of Residential Location" in Spatial Interaction Theory and Planning Models (ed. Karlquist, Lundquist, Snickars and Weibull). Amsterdam: North Holland, 75-96.

Mendelsohn, R. 1980. "The Demand for Characteristics of Goods." Research paper, Department of Economics, University of Washington, 26 pps.

Milon, J. W., J. Gressel, and D. Mulkey. 1985. "Hedonic Amenity Valuation and Functional Form Specification." Land Economics 60:378-387.

Muellbauer, J. 1974. "Household Production Theory, Quality, and the 'Hedonic Technique'" American Economic Review 64:977-994.

Nelson, J. P. 1978. "Residential Choice, Hedonic Prices, and the Demand for Urban Air Quality." Journal of Urban Economics 5:357-369.

Nourse, H. 1967. "The Effect of Air Pollution on House Values." Land Economics 43:181-189.

Palmquist, R. B. 1982. "Estimating the Demand for Air Quality from Property Value Studies." Working paper, North Carolina State, Department of Economics and Business, 44 pps.

Palmquist, R. B. 1983. "Estimating the Demand for Air Quality from Property Value Studies: Further Results." North Carolina State, Department of Economics and Business. Prepared for the Economic Analysis Branch, O.A.Q.P.S.

Palmquist, R. B. 1984. "Estimating the Demand for Characteristics of Housing." Reveiw of Economics and Statistics 64:593-604.

Parsons, G. 1983. "The Identification Problem in Hedonic Price Models." Working paper, University of Wisconsin.

Parsons, G. 1985. "An Almost Ideal Demand System for Housing Attributes." U.S. Environmental Protection Agency.

Polinsky, A. M. and S. Shavell. 1976. "Amenities and Property Values in a Model of an Urban Area." Journal of Public Economics 5:119-129.

Polinsky, A. M. and D. L. Rubinfeld. 1977. "Property Values and the Benefits of Environmental Improvements: Theory and Measurement," in Public Economics and the Quality of Life (L. Wingo and A. Evans, eds.) Washington, D.C.: Resources for the Future.

Pollak [illegible] Household Production Function Approach." Journal of Political Economy 85:1083-1086.

Quigley, J. M. 1982. "Nonlinear Budget Constraints and Consumer Demand: An Application to Public Programs for Residential Housing." Journal of Urban Economics 12:177-201.

Rich, P. R. and L. J. Moffitt. 1982. "Benefits of Pollution Control on Massachusetts' Housatonic River: A Hedonic Pricing Approach." Water Resources Bulletin 18:1033-1037.

Ridker, R. G. and J. A. Henning. 1967. "The Determinants of Residential Property Values with Special Reference to Air Pollution." Review of Economics and Statistics 49:246-257.

Roback, J. 1982. "Wages, Rents, and the Quality of Life." Journal of Political Economy 90:1257-1278.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Price Competition." Journal of Political Economy 82:34-55.

Rothenberg, T. J. 1971. "Identification in Parametric Models." Econometrica 39:577-591.

Rowe, R. D. and L. G. Chestnut. 1982. The Value of Visibility, Cambridge: Abt Books.

Saxonhouse, G. R. 1977. "Regression from Samples Having Different Characteristics." Review of Economics and Statistics 59:124-127.

Schnare, A. and R. Struyk. 1976. "Segmentation in Urban Housing Markets." Journal of Urban Economics 3:146-166.

Schulze, W. D., R. G. Cummings and D. S. Brookshire. 1983. "The Ozone Experiment," Chapter 5 in a report to EPA.

Scotchmer, S. and A. Fisher. 1980. Cross-Cutting Issues in Integrated Environmental Assessment of Energy Alternatives: Evaluation of External Impacts, with Reference to Case Studies. Energy Resources Group, Univesity of California, Berkeley.

Smith, V. K. 1983. "The Role of Site and Job Characteristics in Hedonic Wage Models." Journal of Urban Economics 13:296-321.

Smith, V. K. and T. A. Deyak. 1975. "Measuring the Impact of Air Pollution on Property Values." Journal of Regional Science 15:277-288.

Stevens, J. B. 1966. "Angler Success as a Quality Determinant of Sport Fishery Recreational Values." Transactions of the American Fisheries Society 95:357-362.

Straszheim, M. 1973. "Estimation of the Demand for Urban Housing Services From Household Interview Data." Review of Economics and Statistics 55:1-8.

Straszheim, M. 1975. An Econometric Analysis of the Urban Housing Market. New York: National Bureau of Economic Research.

Theil, H. 1952. "Qualities, Prices and Budget Inquiries." Review of Economic Studies 19:129-147.

van Lierop, J. 1982. "House Price Structure and Market Equilibrium." Journal of Urban Economics 11:272-289.

Wegge, L. L. 1965. "Identifiability Criteria for a System of Equntions as a Whole." Austrialian Journal of Statistics 7:67-77.

Wegge, L. L. and M. Feldman. 1983. "Identifiability Criteria for Muth- [illegible]

Wheaton, W. C. 1977. "A Bid Rent Approach to Housing Demand." Journal of Urban Economics 4:100-217.

Wieand, K. F. 1973. "Air Pollution and Property Values: A Study of the St. Louis Area." Journal of Regional Science 13:91-95.

Willig, R. D. 1978. "Incremental Consumer's Surplus and Hedonic Price Adjustment." Journal of Economic Theory 17:227-253.

Witte, A. D., H. J. Sumka, and H. Erikson. 1979. "An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets." Econometrica 47:1151-1173.