iber

Institute of
Business and
Economic Research

University of
California at
Berkeley

CENTER FOR REAL ESTATE AND URBAN ECONOMICS

WORKING PAPER SERIES

WORKING PAPER NO. 95-235

NONPARAMETRIC HEDONIC HOUSING PRICES

By

Carl Mason
John M. Quigley

WALTER A. HAAS SCHOOL OF BUSINESS

CENTER FOR REAL ESTATE AND URBAN ECONOMICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
Kenneth Rosen, Chair

The Center was established in 1950 to examine in depth a series of major changes and issues involving urban land and real estate markets. The Center is supported by both private contributions from industry sources and by appropriations allocated from the Real Estate Education and Research Fund of the State of California.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
Richard Sutch, Director

The Institute of Business and Economic Research is an organized research unit of the University of California at Berkeley. It exists to promote research in business and economics by University faculty. These working papers are issued to disseminate research results to other scholars. The authors welcome comments; inquiries may be directed to the author in care of the Center.

NONPARAMETRIC HEDONIC HOUSING PRICES

by

Carl Mason

and

John M. Quigley

University of California
Berkeley

WORKING PAPER NO. 95-235

June 1995

Parts of this paper were discussed at the International Conference on Land and Urban Policy, Kyoto, August 1993 and the Seminar on Housing and Housing Markets, Centre for Housing Research, University of Glasgow, March 1994. The paper benefited from the comments of Takahiro Miyao and Christine Whitehead and from seminar participants.

I. Introduction

There is increasing academic and business interest in the measurement of real estate prices and the application of standardized methods to produce price estimates over time and space.

Academic interest arises from the analysis of housing, land, and commercial real estate as economic commodities. For example, economic analysis based on these price estimates has addressed the responsiveness of supply and demand for housing to price variation, the efficiency of the markets for housing and land, and the distributional consequences of the market provision of housing.

Business interest arises from the importance of real estate in the investment portfolios of pension funds and other pools. The efficient allocation of investments by category depends upon the characteristics of returns, that is, upon price trends for real property.

The interests of both academics and practicing professionals have been stimulated by the recent volatility of housing prices and the recognition that individual homeowners are the largest bearers of residential real estate risk. Homeowners are, for the most part, highly leveraged and undiversified. Thus, they have a great deal to gain from the development of derivatives markets in real estate.

These markets would allow individual consumers to hedge the risks associated with their most important investments, their equity in single-family , owner-occupied housing. The development of a futures market for housing requires an accurate and replicable method for measuring the value of housing over time and by region.

The current analytical basis for describing housing price trends is rather weak indeed. In the U.S., for example, the most widely reported measure of housing prices is produced by the National Association of Realtors, and this is merely a compilation of the median sale prices of existing single family homes as reported by member realtors in some 119 U.S. metropolitan areas. The U.S. Bureau of the Census also publishes an index of new single family house prices for each of five large census regions in the country. This index is derived from a regression equation relating the sale prices of houses to a set of independent variables including a couple of size and quality measures. (See Musgrave, 1969, for the original formulation.) Unfortunately this index is not based upon a representative sample of dwellings bought and sold, or even a random sample of new houses completed and sold within a given year. The transactions analyzed by the U.S. Bureau of the Census include only new houses sold in the "speculative builder" category in any year, and thus the index ignores about a third of new houses completed as well as all sales of used housing. (See Peek and Wilcox, 1991, for a discussion.)

It has been widely recognized that it is appropriate to control statistically for the various characteristics of properties in inferring price trends. Indeed, the Bureau of Census index, described above, does so in a crude way. During the last few years, however, there has been increased attention to the statistical problems inherent in the estimation of house price indexes. Much of this attention is addressed to the functional form for the relationship. A variety of reasons have been advanced to indicate why some particular functional form is consistent with economic theory.

This paper demonstrates that this attention is misdirected. Indeed, we prove, using a simple example, that it is a futile exercise to deduce the form of the hedonic relationship from abstract principles of microeconomic theory.

If the form of the hedonic model is purely an empirical matter, it is then appropriate to consider nonparametric procedures. These methods build on the hedonic price framework but relax the constraints on functional form imposed by the traditional methodology. As an example, we present an empirical analysis which compares the nonparametric approach with conventional parametric estimation of housing price indexes. The data for the analysis include an unusual sample of condominium dwellings in the downtown Los Angeles area. The dwellings in the sample are all drawn from a few high-rise buildings located

within a quarter mile of each other. Thus, their locational and public service attributes are quite similar. This provides an excellent opportunity for comparing the methodology of nonparametric estimation with least squares estimation in a situation where regression analysis works well and where the models are straightforward.

II. Hedonic Indexes

A. Parametric Estimation

Hedonic indices of complex commodities have been developed extensively during the past two decades. The original empirical research using this methodology dealt with automobiles (Griliches, 1971) and housing (Kain and Quigley, 1970). These techniques were originally designed to distinguish quality changes over time from price changes for complex commodities with many attributes. In the intervening period, these methods have been widely diffused, especially in housing market research.

Hedonic techniques typically result in the estimation of some regression relationship between the sale price or monthly rent of individual properties, V_t, their physical and locational characteristics, x, and some specification of time, t

$$(1) \quad V_t = f(x,t) \quad .$$

The appropriate interpretation of this relationship depends crucially upon the inclusion of the correct set of the property characteristics, x, and the correct functional form, f(.), for the hedonic regression. Conditional upon choice of the appropriate variables and functional form, the hedonic function can be used to disaggregate the intertemporal variation in prices into that attributable to changes in the qualitative and quantitative characteristics of properties sold and that attributable to intertemporal variation in the unit prices of these characteristics. Importantly, the statistical results can be used to produce indices of the market price for a "standardized" or quality-adjusted property over time or space.

These two maintained hypothesis (the "correct" set of x and the "correct" functional form f[.]) are not subject to simple tests. The problem of selecting the correct set of independent variables in a regression model is familiar in applied economics. But it is not easily solved, at least not by simple rules. The problem of selecting the correct functional form is also long-standing, but it has only recently become a popular area of research.

It has often been asserted that the appropriate functional form for the hedonic regression of housing and real estate prices can be deduced from economic theory (see Colwell, 1993, for a recent statement). This is incorrect. The hedonic relation can, of course, be derived from the underlying microeconomics of consumer choice. But this does

not mean that the form or the curvature of the relationship can be determined from abstract principles alone.

To demonstrate this, we present a simple example, based on rudimentary microeconomic principles which indicates how the hedonic function is derived from the behavior of consumers. We then place some simple but plausible restrictions on consumer preferences and show that they imply practically nothing at all about the hedonic house price function.

A Simple Example: Assume that consumers of income y derive utility from housing x (a one dimensional commodity) and other goods z at a price of one ($P_z = 1$). Housing is sold to the highest bidder. Assume that the stock of housing is fixed and that households vary only in their incomes. Then, if housing is a normal good, competition will force the household with the lowest income to choose the smallest or lowest quality dwelling in the market. Because housing is a normal good, the function $y(x) = G(x)$, which relates a household's income to its housing consumption, will be monotonically increasing over the range of housing in the market. The G(x) function can be thought of as the outcome of a hypothetical auction in which each household submits bids for each available house, and the highest bidder for each house wins. Unlike the standard consumer maximization problem in which the household may purchase any quantity of x at a given price, the market clearing hedonic price of *each* house arises endogeneously

from the competition of consumers. The nature of the price variation is determined by the interaction of the given housing supply and the distribution of income across households. The market clearing "hedonic" price function is determined, *not only* by preferences of households *but also* by supply conditions and by the distribution of income as well.

The hedonic price function V(x) is derived directly from the conditions for maximization of consumer utility and the mapping, y = G(x).

Let U(x,z) represent the consumers' utility functions. The conditions for utility maximization insure that the ratio of the marginal utility of x to that of z equals the ratio of their marginal prices. Thus,

$$(2) \qquad \frac{dV(x)}{dx} = \frac{\partial U(x,z)/\partial x}{\partial U(x,z)/\partial z}$$

$$= \frac{\partial U(x,y-V[x])/\partial x}{\partial U(x,y-V[x]/\partial z}$$

$$= \frac{\partial U(x,G[x]-V[x])/\partial x}{\partial U(x,G[x]-V[x]/\partial z}$$

The term on the left hand side is the marginal price of housing, that is, it is the derivative of the hedonic price

function. The term on the right hand side is the marginal rate of substitution of housing for other goods. Substituting G(x) for y in the last line yields a differential equation in x alone. The hedonic price function, V(x) is the solution to this differential equation (with some initial condition $V(x_0) = C$). Since G(x) is the equilibrium allocation of housing, the V(x) which maintains that allocation and also satisfies equation (2), the first order conditions of all consumers, is a stable pareto optimal hedonic price function.

Note that a wide variety of hedonic functions with very different properties can be specified by equation (2) even for a specific utility function, depending upon the distribution of income and housing in the local market (that is, the shape of the y(x) = G(x) function).

The function y = G(x) reflects the distribution of income and the distribution of housing attributes in the local market. Ceteris paribus, if the distribution of incomes is more equal, then G(x) will be more steeply sloped. Small increases in income will be associated with larger increases in housing consumption. Similarly, if there is more variation in the quality of housing, then G(x) will be flatter. Thus, the peculiarities of local housing markets will affect the functional form of the hedonic function *even if* consumers in all markets have the same utility functions and *even if* the income distribution in all markets is the same.

For specific example, let $G(x) = x^{\delta}$, and let the utility function be Cobb-Douglas, $U = x^{\alpha}z^{\beta}$. The differential equation describing the hedonic function is

$$(3) \quad \frac{dV(x)}{dx} = \alpha[x^{\delta} - V(x)]/\beta x$$

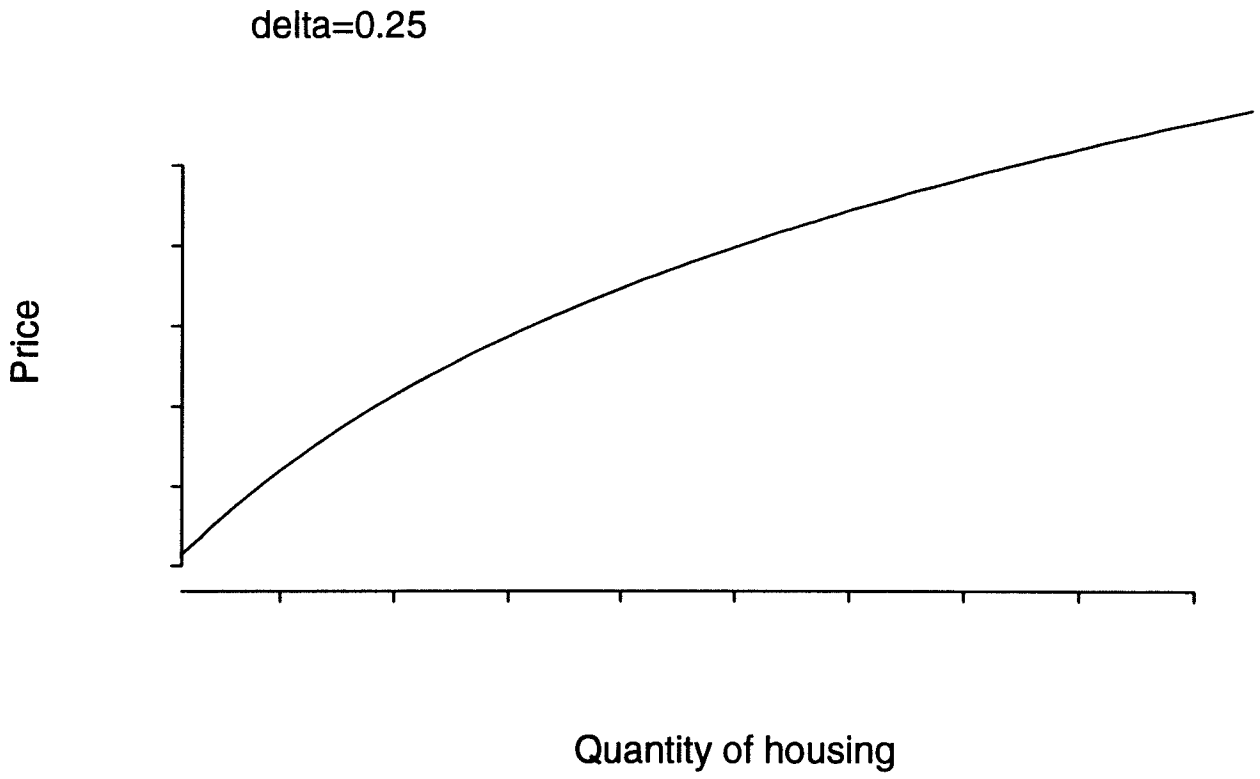
With $V(1) = 1$ as the initial condition, the solution of (3) is,

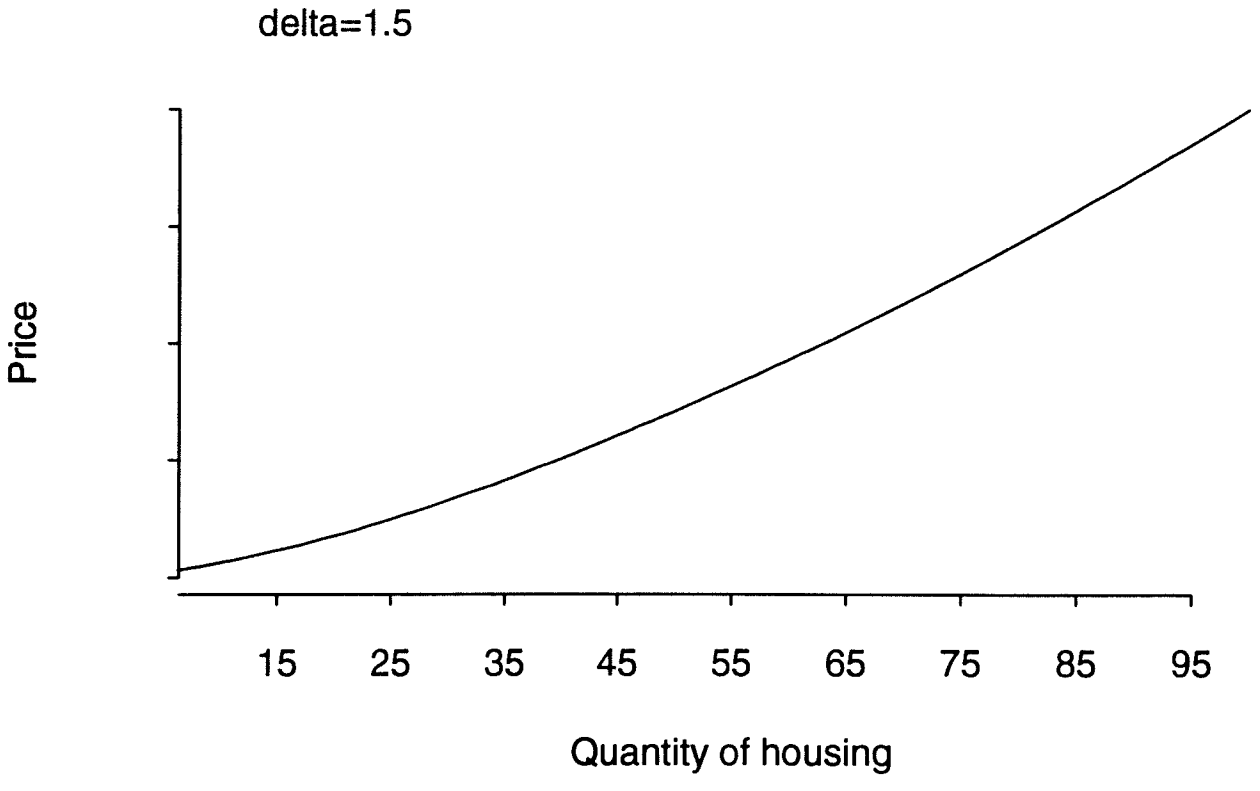
$$(4) \quad V(x) = [\alpha x^{\delta} + \delta\beta x^{-\alpha/\beta}]/[\alpha + \delta\beta]$$

Even in this very simple case, the shape of the hedonic price function is clearly sensitive to changes in $G(x)$. Figure 1 graphs the hedonic price function for two values of δ. As the figure indicates, the price function is increasing for both values of δ -- dwellings containing more or better housing are more expensive. However, one function is concave and the other is convex. Note that the consumers' utility functions which underly the two curves are identical. (In both cases the utility function is Cobb-Douglas with $\alpha = \beta = 0.5$). The only difference is in the $G(x)$ function. The dotted line is drawn with $G(x) = x^{1/4}$. The solid line is based on $G(x) = x^{3/2}$. Clearly theory alone tells us practically nothing about the shape of the hedonic price function.

Figure 1

Hedonic price function for different values of delta





The inability to specify the functional form of the hedonic index has led to suggestions that the characteristics of properties be standardized with reference only to themselves (e.g., Case and Shiller, 1987), and that the effect of intertemporal variation be analyzed by relying on some very general specification. Standardizing properties with reference to themselves implies an analysis based upon repeat sales of properties whose characteristics have remained unchanged between sales. However, the benefits of avoiding specifying and measuring housing characteristics using this approach come at considerable cost.[1]

B. Non-Parametric Estimation

Consider, as an alternative, the estimation of (1) by a non-parametric method, for example the Generalized Additive Model (GAM). The GAM represents a compromise between generality and comprehensibility. At the extreme of comprehensibility are conventional (parametric) hedonic regression techniques. At the extreme of generality are techniques such as local or kernel regression in which smoothed surfaces are generated from neighborhoods of points. While these latter procedures impose very few assumptions, they become very difficult to interpret if the set of independent variables is even of moderate size.

---

[1] These costs are discussed in detail in Quigley (1995).

One paper has applied general smoothing procedures to estimate price indices for residential housing. Meese and Wallace (1991) used the technique of local regression to estimate housing price indexes for fourteen California communities during the 1970-1988 period. By using nonparametric methods, they were able to avoid assuming the same functional form for each submarket. But the high dimensionality of the data made it difficult to produce understandable results. Meese and Wallace dealt with this interpretation problem by constructing Fisher price indexes. While this is perfectly workable, it adds an additional level of complexity to the estimation procedure.[2]

The alternative method we explore in this empirical analysis is a general additive model. In place of the completely general techniques of Meese and Wallace, we impose the assumption of additivity of effects. The cost of this assumption is the suppression of complex interactions, but the benefits in terms of tractability of the results are significant. (See Hastie and Tibishirani, 1990).

Consider the relationship,

$$(5) \qquad V_{it} = \sum_j S_j(x_{ij}) + S_o(t) + e_i \quad ,$$

where V_{it} is the rent or sale price of dwelling i at time t, x_{ij} is housing characteristic j measured for observation i

[2] It also creates the need for complex boot strap procedures for estimation of standard errors.

and $S_j(.)$ is an arbitrary smoother applied to the jth variable. Assume that $E(e_i)=0$ and $E(e_i^2)=\sigma^2$. Because of the additive structure, the smooth function $S_j(x_{ij})$ captures the entire effect attributable to x_{ij}. Because $S_j(.)$ is of low dimension, each of the $S_j(.)$ can be easily displayed and interpreted visually. Because $S_j(.)$ is a function of only one (or a few) variables, it is analogous to a coefficient estimated by ordinary least squares in a linear additive hedonic model.

In a practical situation, the GAM can be represented as

$$
\begin{aligned}
(6) \quad s_o &= \mathbf{s}_o\,(V - * - s_1 - s_2 - \ldots - s_j) \\
s_1 &= \mathbf{s}_1\,(V - s_o - * - s_2 - \ldots - s_j) \\
s_2 &= \mathbf{s}_2\,(V - s_o - s_1 - * - \ldots - s_j) \\
&\vdots \quad\quad \vdots \\
s_j &= \mathbf{s}_j\,(V - s_o - s_1 - s_2 - \ldots - *)
\end{aligned}
$$

where the asterisks are place holders showing the term that is missing in each row.

Here lower case s represents a vector of values which are the smooth function S evaluated at each observation. In each case the argument of the smooth function is V minus all of the other smoothers similarly evaluated. Equation (6) is a simultaneous system of j+1 equations. Iterative methods such as the Gauss-Seidel algorithm (see Chambers and Hastie, 1992) can solve such a system efficiently. The algorithm

simply recalculates each smoother at each cycle using the current values for all of the other smoothers.[3]

For a broad class of smoother functions S, estimates can be shown to be consistent, and exact and approximate standard errors of the estimates can be constructed (See Hastie and Tibishirani, 1990).

III. Empirical Analysis

The empirical analysis is based upon a sample of 843 condominium sales recorded during the twelve year period from January 1980 through December 1991 in downtown Los Angeles. The sample includes essentially every condominium sale within the downtown area of Los Angeles during this time period. Condominiums were located in four different high rise properties which realtors and real estate agents consider "comparables" for the purpose of appraisal. There are no other "comparables" within several miles of downtown. We gathered information on the original sellout prices of each of the condominiums in one of the high rise properties completed in 1989 and all subsequent sales of these dwelling units. We also obtained information on all condominium sales in each of the other three properties beginning in

[3] In the special case where each of the S_j is the least squares projection, the equation system in (6) reduces to the usual normal equations for ordinary least squares. One recent paper applies a special case of (6) to the analysis of real estate markets (Coulson, 1992). In this application one variable (floor area) is represented by a general smoother and the others are least squares coefficients.

1980. Property characteristics were obtained by matching addresses to condominium floor plans. Resale information was obtained from multiple listing services, from court records and from real estate lenders.

Because the sample consists of properties in only four high rise buildings located within about a quarter mile of each other, the neighborhood and public service amenities associated with these properties are identical. The condominiums vary in their size and their location within each of the buildings. We recorded the date of each sale and the selling price of the property.[4] In none of the condominiums, were the physical characteristics of the sale properties changed during the sample period.

The statistical analysis reported below relates the selling prices of these apartments to the sizes and locations of the properties, the projects in which they are located and the timing of sales.

Due to the unusual nature of the data set, the problems associated with choosing the correct set of independent variables, x, are not particularly serious. Among all of the properties in the sample, there are a limited number of floor plans and locations. Consequently, characteristics such as floor area are not likely to be seriously distorted by unmeasured features. Further, since the properties are condominiums, significant changes in physical

[4] Selling prices are reported in real terms, using the quarterly Consumer Price Index (See **Economic Report of the President**, 1992).

characteristics are practically impossible. This makes it possible to concentrate on the effects of relaxing the assumptions about functional form, rather than on the difficulties associated with selecting the correct set of covariates. We report the results of estimation using standard hedonic regression and estimation using the (nonparametric) generalized additive model.

A. Parametric Hedonic Price Indexes: The Standard Model

We estimate equation (1), in the semi-log specification, in three variants using different representations of the timing of sales. The simplest representation merely includes the date of the sale, t, as an independent variable.

$$(7) \quad \log V_t = a + \sum_{i=1}^{5} b_i x_i + ct$$

There are five control variables, x's: size (in square feet) and location (story) and dummy variables for each of three buildings. In this formulation the time of the sale, t, is measured by the number of days since January 1, 1980.

In the second formulation, we use the same set of control variables but we replace the simple time trend with a set of dummy variables measuring quarter year of sale, beginning January 1, 1980.

In the third formulation, we represent time by a set of dummy variables for year of sale, again beginning with 1980.

Regression estimates of the three variants of the hedonic index are summarized in Table 1. Condominium prices vary substantially with the size of the dwellings and their locations--larger dwellings on higher floors command a premium. They also vary significantly with the specific building--projects B and C are more desirable than the other projects. The coefficients are stable across the three specifications.

The real selling price of otherwise identical buildings clearly varies over time. In the simplest specification, a linear time trend has a t ratio of about 20. This model explains 73 percent of the variation in selling prices. The most complex specification, where time is represented quarterly, explains an additional 3 percent of the variation in housing prices.

B. Nonparametric Estimation

The nonparametric (or semi parametric) model which corresponds to equation (7) is

$$(8) \quad \log V_t = S_1(x_1) + S_2(x_2) + \sum_{i=3}^{5} b_i x_i + S_o(t)$$

where $S_1(.)$, $S_2(.)$ and $S_o(.)$ are arbitrary smoothers applied to the variables measuring size, location, and time

TABLE 1

Parametric Standard Hedonic Price Indices For Downtown Los Angeles

Dependent Varible: Logarithm of Selling Price

	Model 1 Linear Time Trend	Model 2 Time in Years	Model 3 Time in Quarter Years
Size (1000 sq. ft.)	0.867 (38.39)	0.862 (38.40)	0.876 (38.76)
Location (story)	0.096 (7.94)	0.095 (7.83)	0.096 (7.94)
Project A (dummy)	-0.061 (2.01)	-0.073 (2.23)	-0.078 (2.30)
Project B (dummy)	0.205 (7.17)	0.201 (6.80)	0.199 (6.53)
Project C (dummy)	0.138 (6.91)	0.113 (4.63)	0.099 (3.96)
Intercept	14.042 (207.18)	13.732 (105.33)	14.274 (126.27)
Time (1000 days)	-0.118 (20.32)	*	**
R^2	0.727	0.738	0.759

Notes:

Each regression is based on 843 observations on condominium sales.

* Regression also includes 11 variables measuring time in years.

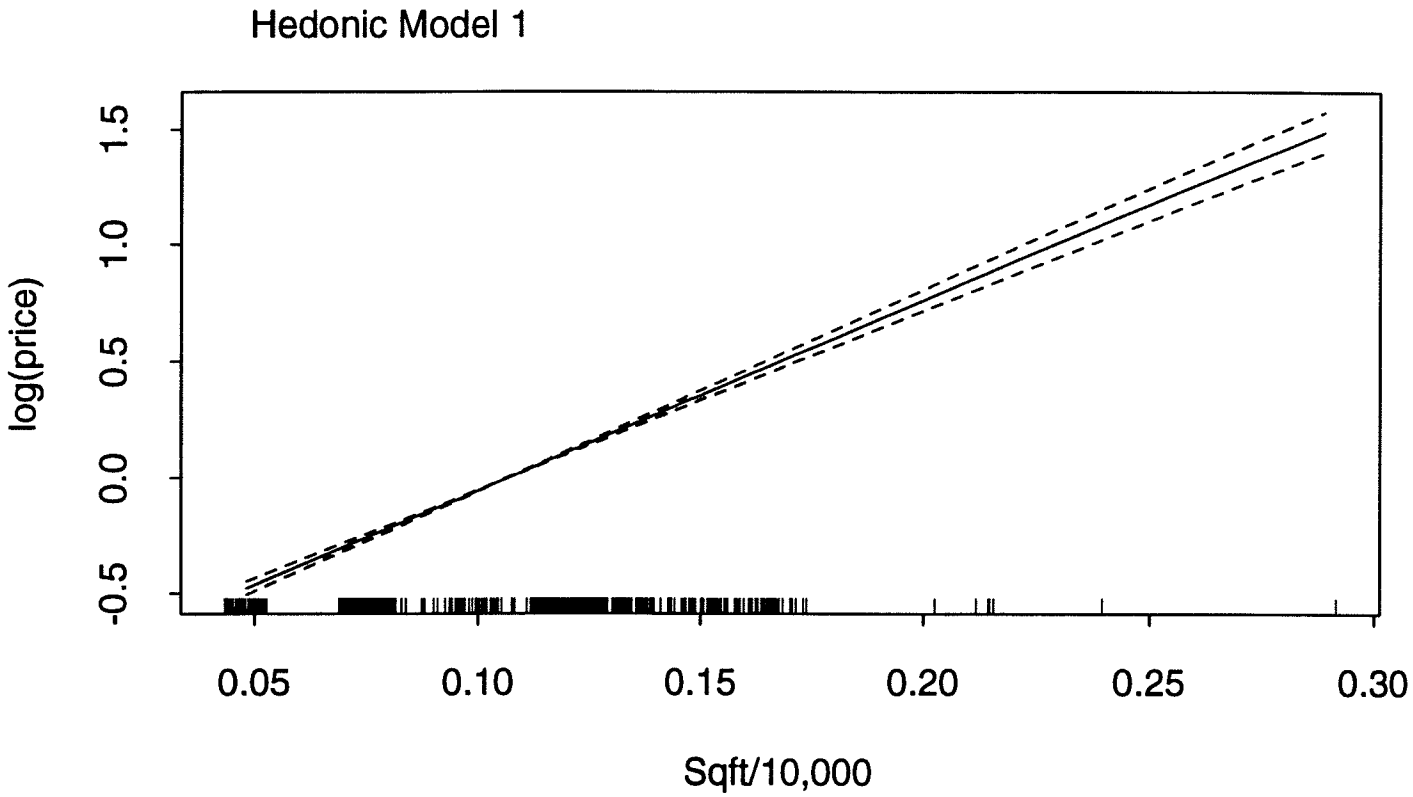
** Regression also includes 47 variables measuring time in quarters.

respectively. The functions are estimated iteratively applying equation (6) using standard computer software (as described in Chambers and Hastie, 1992). In the second formulation, we replace the time trend $S_0(t)$ with a set of dummy variables measuring quarter of sale, and in the third formulation we represent time by dummy variables measuring sales during each quarter of the period.

For each of the variants, the estimated smooth functions are highly significant, with probability levels of 0.003, 0.060, and 0.007 respectively. Figures 2, 3, and 4 illustrate the differences between the hedonic model and the GAM estimator. Each comparison is based upon Model 1. The solid lines represent the estimated relationship, while the dotted lines represent the 95 percent confidence interval. Figure 2 reports the percent change in price (i.e., the logarithm of sale price) as a function of the size of the dwellings. As compared to the linear relationship imposed by the hedonic function, the more general GAM relationship reveals a slightly less than proportional relation between the size of dwellings and the percent change in housing prices. Figure 3 presents the same comparison for the variable indicating the floor (story) in which the property is located. In this comparison, the relationship estimated by the GAM is not quite linear, but it is very close. Figure 4 presents a comparison of the time trend estimated by imposing a linear relationship and that estimated using the more general GAM technique. As the GAM result

Figure 2

Comparison of the effect of size on Los Angeles condominium prices



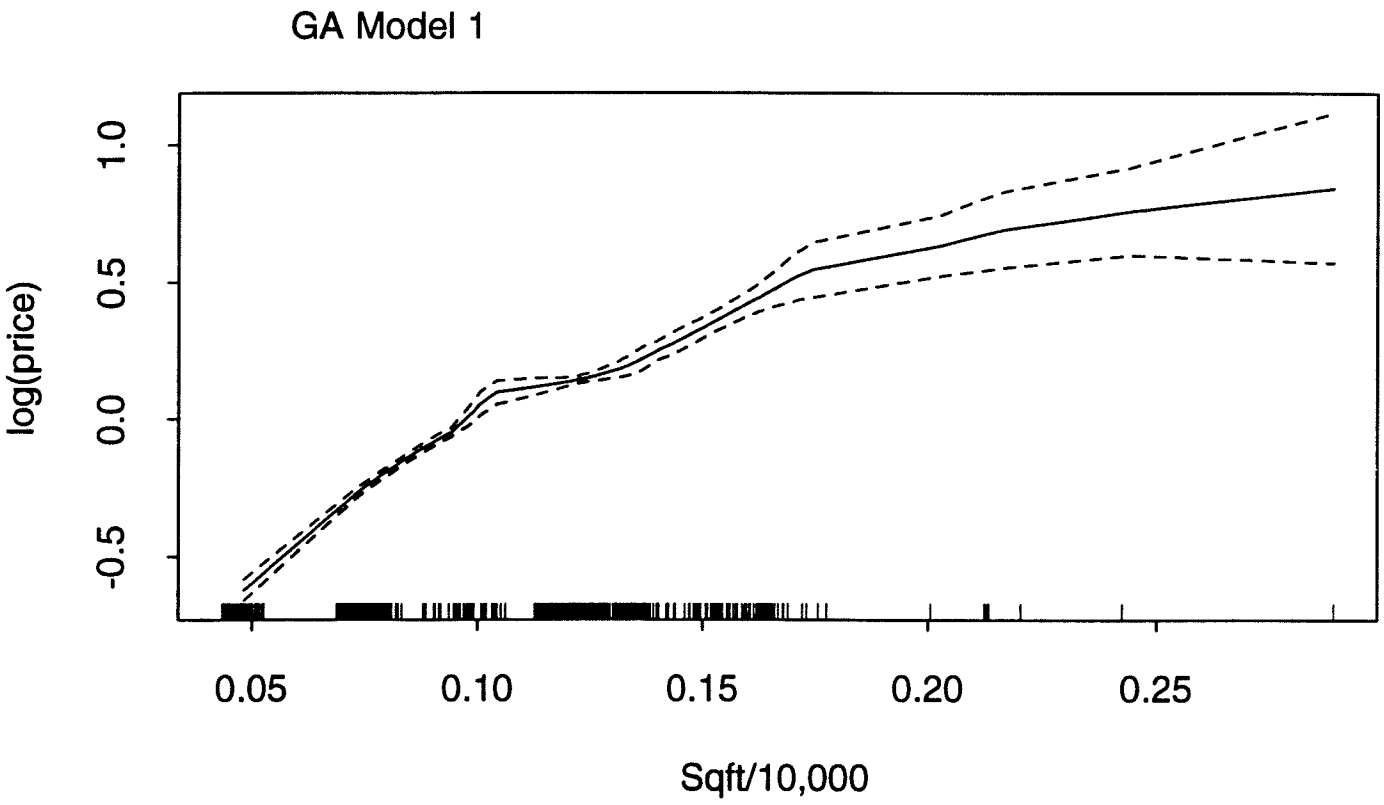
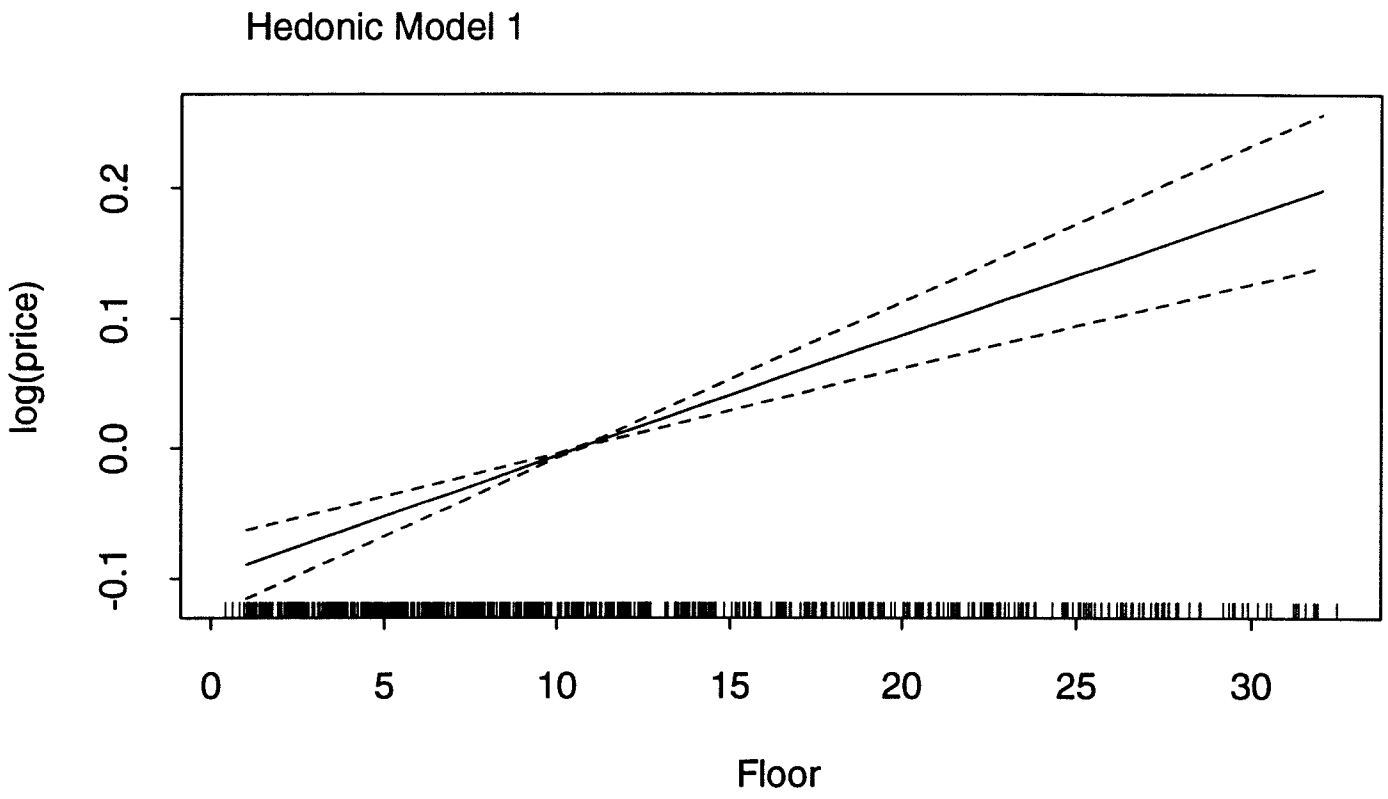


Figure 3

Comparison of the effect of location on Los Angeles condominium prices



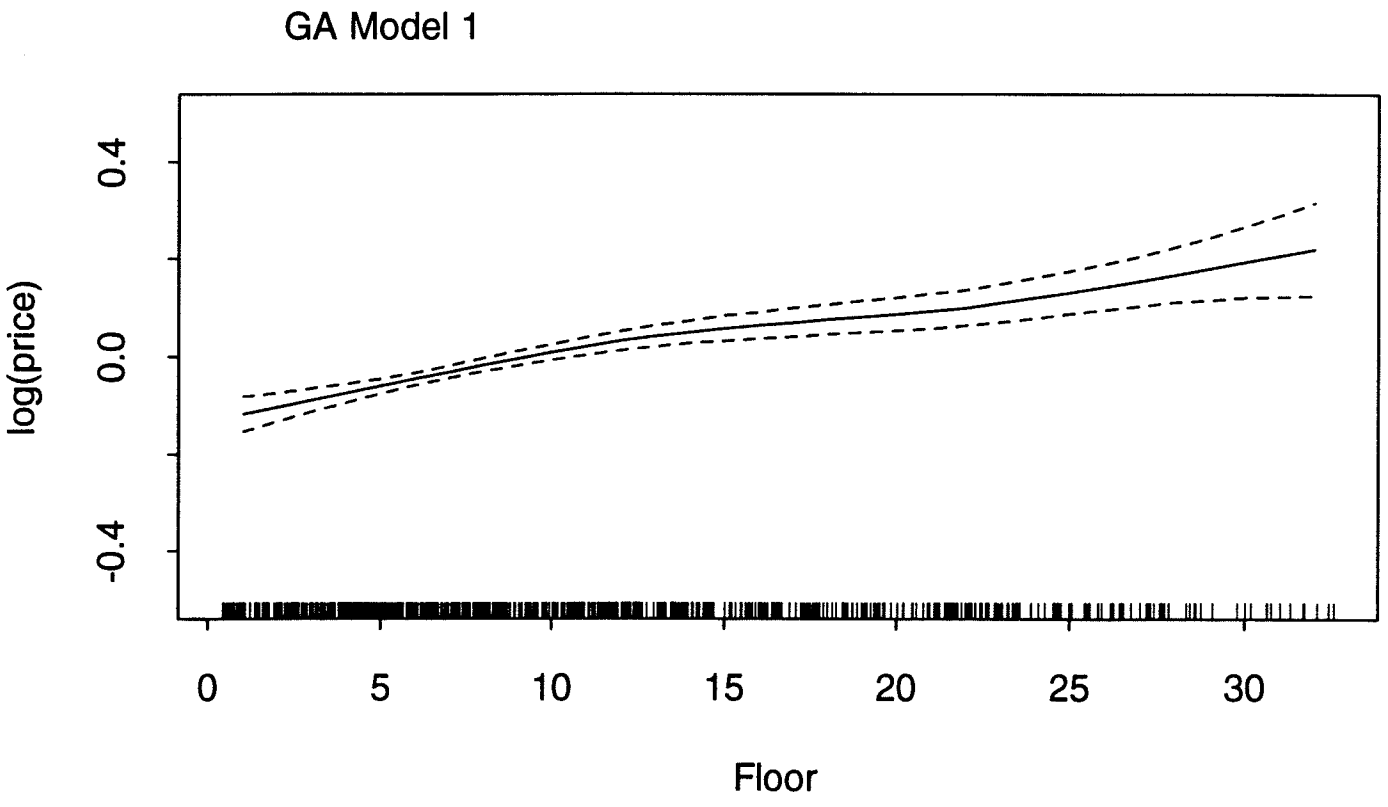
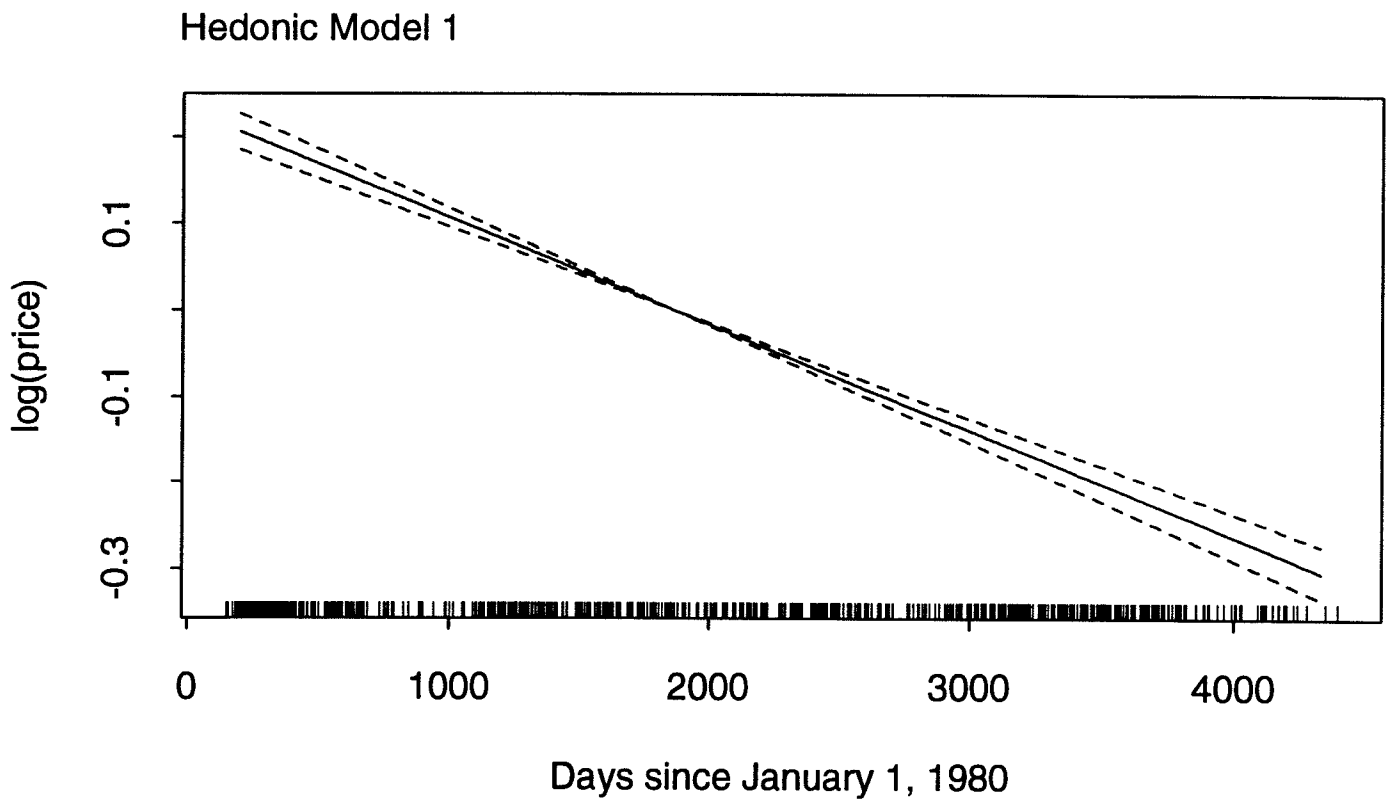
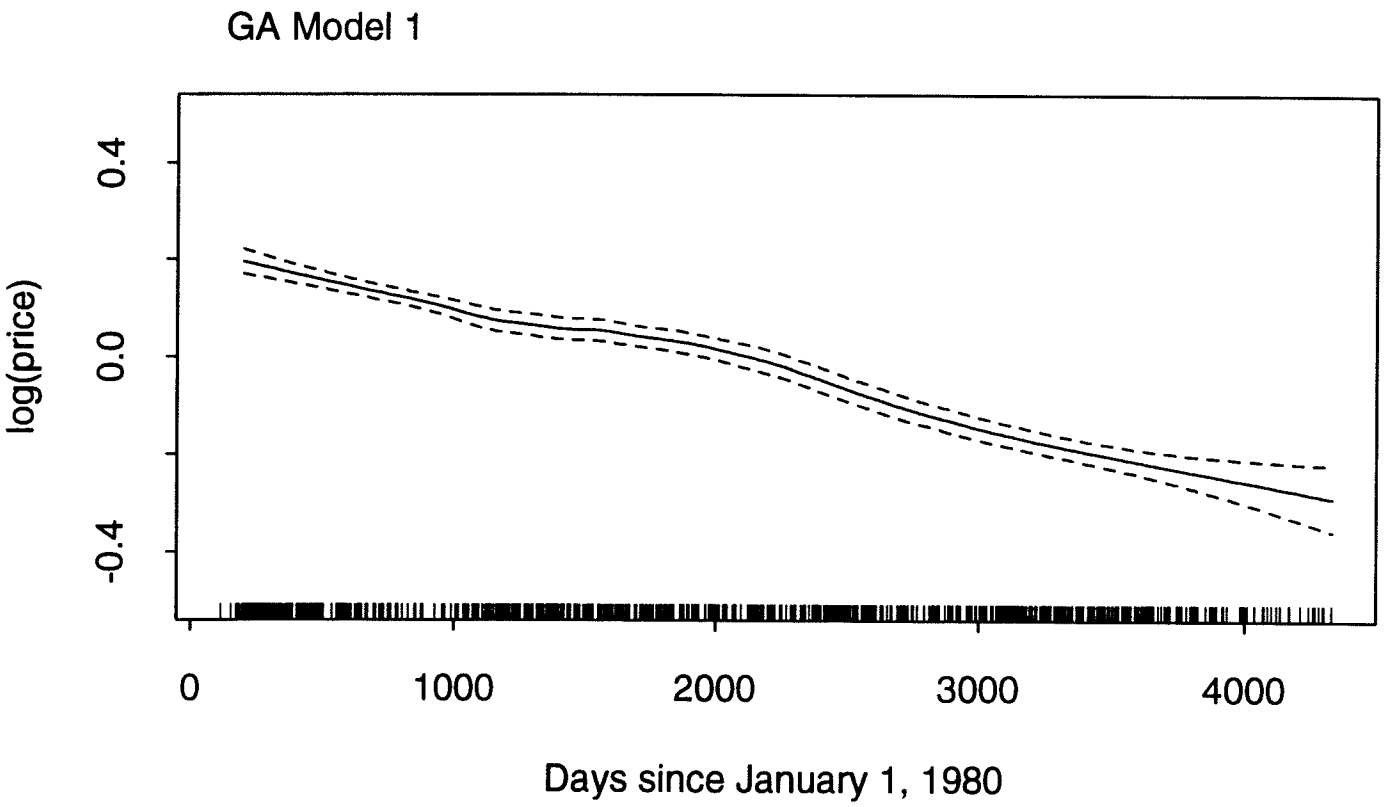


Figure 4

Comparison of time trends in Los Angeles condominium prices





indicates, the effect of time is not quite linear during the 1980-1991 time period.

In each case, the GAM representation is significantly different, in a statistical sense, from the simple linear relationship, but none of the differences reported is very large.

IV. Conclusion

As the accurate measurement of housing price movements has become more important, it is natural that increased attention be paid to the theory underlying hedonic prices and the methods used to estimate them. In this paper we demonstrate that the theory of consumer behavior does not provide guidance about the choice of the form of the relationship.

Since the form of the relationship is an empirical matter, we suggest that estimation of the hedonic function by non parametric methods offers potential advantages. The GAM method is quite tractable. It is straightforward to compute and is easily interpreted. Importantly, the GAM estimator imposes no functional form **a priori** upon the hedonic relation.

A comparison of models estimated using data on condominium sales in downtown Los Angeles during the decade of the 1980's is suggestive of the advantages of GAM estimation. The GAM estimates relax the assumptions of linearity imposed by regression estimation. Graphical

comparisons of the GAM estimator with the linear regression estimator confirm a non linear relationship between housing attributes and price. In this application, however, the departures from linearity are rather small, though statistically significant.

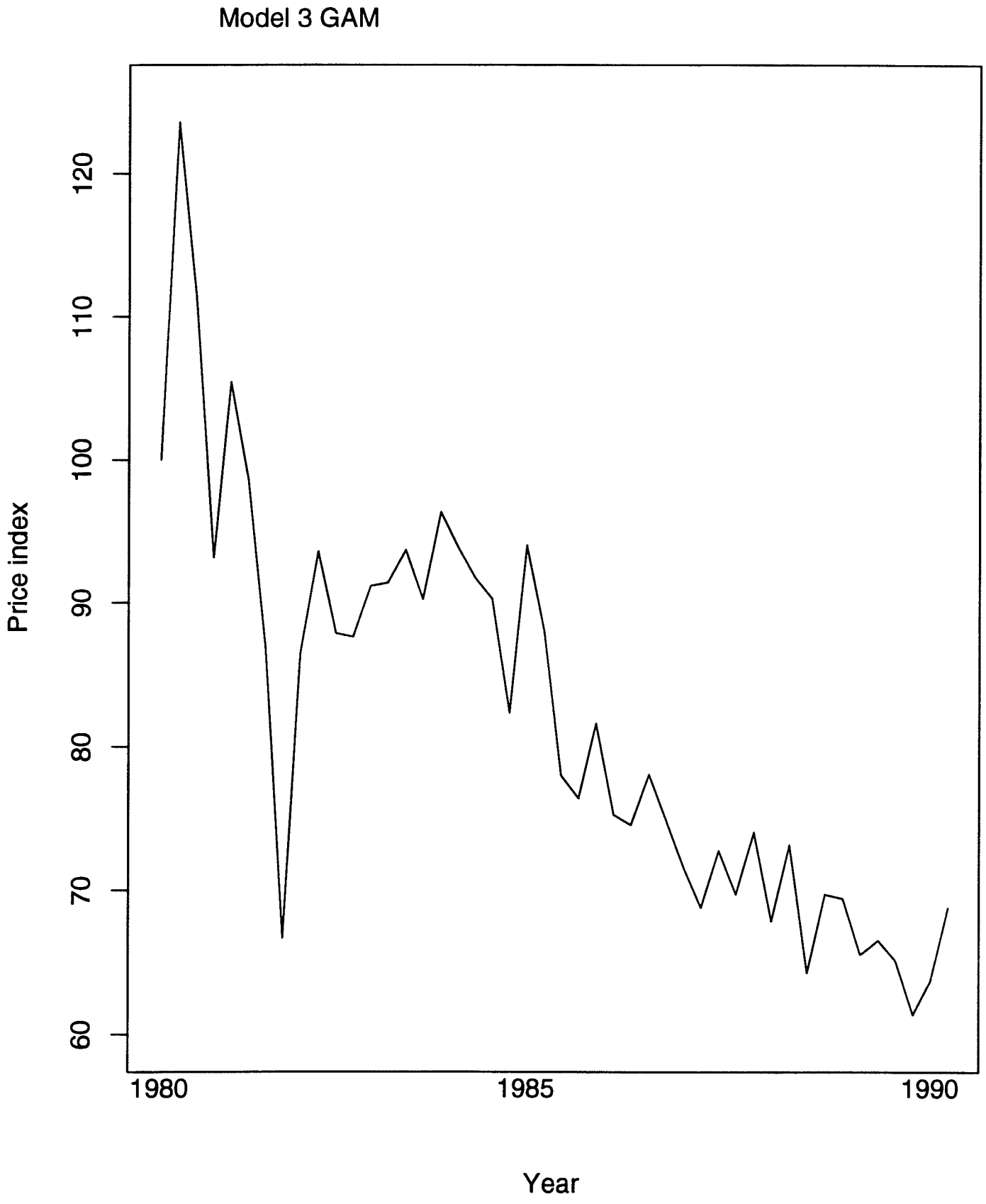
Substantively, the analysis of housing prices in downtown Los Angeles documents the collapse of the Southern California housing market during the decade of the 1980's. Figure 4, in particular shows the decline in housing prices.

As argued in this paper, the most credible estimates of the trend in housing prices during this period are provided by the Generalized Additive Model. Figure 5 presents the quarterly time trend estimated from Model 3 using this method of estimation. As the figure shows, the decline in house prices during the 1980-1991 period was spectacular. The results of Model 3 suggest that downtown condominiums in Los Angeles lost about forty percent of their value during the decade.

Figure 5

Quarterly index of condominium prices in Los Angeles

1980:I = 100



References

Case, Karl E. and Robert J. Shiller, "Prices of Single Family Homes Since 1970: New Indexes for Four Cities," **New England Economic Review**, September/October, 1987: 45-56.

Chambers, John M. and Trevor J. Hastie, **Statistical Models in S**, Pacific Grove, CA: Wadsworth and Brooks, 1992.

Colwell, Peter F., "Comment: Semiparametric Estimate of the Marginal Price of Floorspace," **Journal of Real Estate Finance and Economics** 7(1), 1993: 73-76.

Coulson, N. Edward, "Semiparametric Estimates of the Marginal Price of Floorspace," **Journal of Real Estate Finance and Economics** 5(1), 1992: 73-84.

Griliches, Zvi, ed., **Price Indices and Quality Change**, Cambridge MA: Harvard University Press, 1971.

Hastie, T. and R. Tibshirani, **Generalized Additive Models**, London: Chapman and Hall, 1990.

Kain, John F. and John M. Quigley, "Measuring the Value of Housing Quality," **Journal of The American Statistical Association**, 65, 1970: 532-548.

Mason, Carl and John M. Quigley, "Comparing the Performance of Discrete Choice and Hedonic Models," in Manfred M. Fischer, et al, eds., **Spatial Choices and Processes**, Amsterdam: North Holland, 1990.

Meese, Richard and Nancy Wallace, "Non-Parametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices," **The AREUEA Journal** 19(3), 1991: 308-333.

Musgrave, John C., "The Measurement of Price Changes in Construction," **Journal of the American Statistical Association** 64(237), 1969: 771-786.

Peek, Joe and James A. Wilcox, "The Measurement and Determinants of single Family House Prices," **The AREUEA Journal**, 19(3), 1991: 353-382.

Quigley, John M., "A Simple Hybrid Model for Estimating Real Estate Price Indices, " **Journal of Housing Economics**, 4(1), 1995: 1-12.