Combining Revealed and Stated Data to Examine Decisions of Housing Location:

Discrete-Choice Hedonic and Conjoint Analysis

Dietrich Earnhart
University of Kansas

Department of Economics
University of Kansas
213 Summerfield Hall
Lawrence, KS 66045
(785) 864-2866
Earnhart@ukans.edu

June 30, 1998

Abstract: This paper uses stated preference and revealed preference data, separately and jointly, to examine individuals' choices regarding housing locations. In particular, it combines an established revealed preference approach, discrete-choice hedonic analysis, and a relatively new stated preference approach, choice-based conjoint analysis, to understand better individuals' housing decisions. These methods are appropriately combined since they reflect the same decision process: household selects the location that provides the best combination of attributes from the feasible set of alternatives. Analysis finds that actual and hypothetical housing purchases are similar decision processes with respect to some attributes, yet dissimilar with respect to other attributes. Nevertheless, combining the two types of data improves identification of both attribute categories.

JEL Classification Number: R21

I thank Joffre Swait for his indispensable guidance, encouragement, and technical assistance. I thank David Popp, Joshua Rosenbloom, Don Lien, and Tom Weiss for their helpful suggestions. I thank Tom Stanke, director of the Town of Fairfield Conservation Commission, for his steadfast support. I thank Claire Walden, Zac Graves, and Wenli Huang for their excellent research assistance. Finally, I acknowledge the financial assistance of Fairfield University who provided me with a research grant. All errors remain my own.

1. Introduction

Numerous empirical studies seek to estimate the demand for housing. Many studies employ a hedonic framework to examine individuals' choices of housing locations. Most of these studies use the hedonic price framework articulated by Rosen (1974), which assumes that a continuous function relates the price of a house to its attributes — the hedonic price function — and that people select a house by equating the marginal utility of each house attribute to its marginal price. The prominent examples are Cropper et. al. (1988), Graves et. al. (1988), Palmquist (1984), Bartik (1987), and Brown and Pollakowski (1977). However, the decision to buy a house is more naturally framed within a discrete-choice framework of hedonic analysis: a household chooses one dwelling from a large set of discrete and heterogeneous alternatives (Quigley, 1985). In addition, the discrete-choice framework does not need to assume the existence of a market equilibrium in order to generate results, unlike the hedonic price framework. Nevertheless, relatively few empirical studies have employed this framework since McFadden (1978) first theoretically articulated it. The prominent studies are Quigley (1976), Williams (1979), Friedman (1981), Longley (1984), Quigley (1985), and Nechyba and Strauss (1998).[1]

While discrete-choice hedonic analysis examines the housing decision process by observing how individuals select the housing location that provides the best combination of attributes, choice-based conjoint analysis attempts to mimic this selection by asking respondents to identify their choice from a hypothetical set of housing locations. Each set is generated by varying housing location

[1] Ellickson (1977), Ellickson (1981), and Gross et. al. (1990) apply a discrete-choice model to housing choices using a probabilistic choice bid-rent framework, which is different from McFadden (1978)'s framework. In addition to housing choices, other studies use the discrete-choice framework to examine tenure and transportation choices (Lerman, 1977; Boehm et. al., 1991; King, 1980; Anas and Chu, 1984; Anas, 1982).

attributes. In other words, each housing alternative is defined by its attributes. Despite the usefulness of this approach, very few studies use choice-based conjoint analysis to examine housing decisions (Timmermans and van Noortwijk, 1995; Timmermans et. al., 1992).

This paper combines these two useful approaches, discrete-choice hedonic analysis and choice-based conjoint analysis, to understand better housing decisions. The combination is straightforward given the similar construction of the two models — both reflect the same decision process of selecting a housing location. In both cases, individuals select one housing location from all locations available to them, where the selection of housing location is modeled as a function of price and other attributes of the location. Discrete choice random utility theory and multinomial logit estimation techniques apply to both models and generate comparable measurements.

Each of the individual analytical methods has its strengths and weaknesses (which the next section describes in detail). By combining the discrete-choice hedonic and conjoint analysis, the joint model enhances the strengths and diminishes the weaknesses of each individual method. This combination yields two important benefits. First, the combined approach should generate an econometric model with greater explanatory power, more robust parameter estimates, and improved identification of influential parameters. Of course, joint analysis may not indicate similar decision processes but instead reveal differences between the revealed and stated decision processes. If so, the combined approach can indicate which parameters of the decision process are similar and which cause actual behavior to differ from stated intentions. Even in this case, the combined approach should improve the identification of influential parameters.

No previous study combines (or compares) discrete-choice stated and revealed preference models to examine housing choices. Moreover, no previous study combines any two stated and revealed preference models to explore such choices. The only similar analysis is Goodman's (1989), which links estimation results from factorial survey analysis, a stated preference method, and estimation results from hedonic price analysis to value structural and neighborhood attributes of housing. In research areas other than housing, recent studies combine stated and revealed preference methods to understand individual's economic decisions regarding recreation, transportation, and durable goods (Cameron, 1992; Chapman et. al., 1996; Adamowicz et. al., 1994; Adamowicz et. al., 1997; Huang et. al., 1997; Adamowicz and Swait, 1996; Train and Atherton; Swait et. al., 1994).[2]

To examine the combination of revealed and stated data, this research uses data on actual housing location choices made by individual households living in Fairfield, CT, and data on hypothetical housing location choices generated by distributing mail surveys to the same group of individuals. This approach is more helpful than linking two different groups of homeowners and survey respondents as Goodman (1989) does.

The remainder of the paper details these points. Section 2 describes the full rationale for combining these stated preference and revealed preference methods. Section 3 formulates the theoretical framework. Section 4 depicts the analytical approaches for data collection. Section 5 structures and interprets the econometric analysis. Section 6 summarizes.

2. Rationale for Combining Hedonic and Conjoint Analysis

Previous research utilizes the hedonic and conjoint analytical methods to examine household

---

[2] Cameron (1992), Chapman et. al. (1996), and Huang et. al. (1997) combine the continuous choice-based travel cost and contingent behavior methods. Adamowicz et. al. (1994), Swait and Adamowicz (1996), and Adamowicz et. al. (1997) combine the discrete choice-based travel cost and conjoint methods. Train and Atherton (1995) combine stated and revealed preference data to examine customers' choices of appliance efficiency level. Swait et. al. (1994) examine businesses' choices of freight shipping alternatives.

choices of residential location. Numerous studies use hedonic analysis to examine housing markets. Most studies apply the hedonic price model, which assumes people select a house by equating the marginal utility of each house attribute to its marginal price (Rosen, 1974). Relatively few previous studies apply the discrete choice hedonic model, which views the individual as choosing the house that gives him/her the highest utility from all the houses in a universal choice set, with utility a function of attributes (McFadden, 1978).[3] In order to combine the revealed and stated methods within a common theoretical framework, this paper employs the discrete-choice hedonic model.

In the economics literature, conjoint analysis takes different forms. Rank-ordered conjoint analysis (also called factorial survey or vignette analysis) produces descriptions of various “goods” and asks respondents to rank or rate the goods; Goodman (1989) uses this method to examine housing choices.[4] This approach seems inappropriate for explaining housing purchases since it does not mimic the actual behavior of house buyers; although buyers may rank houses initially, the most relevant decision is the purchase of a single home (Freeman, 1991). Instead, choice-based conjoint analysis is more appropriate since it asks respondents to choose one housing location from a set of constructed housing alternatives. While numerous studies use this form of conjoint analysis to analyze the demand for common market goods (Bunch et. al., 1992; Louviere and Hensher, 1982; Louviere and Woodworth, 1983), only two studies apply the choice-based version to the market good of housing (Timmermans and van Noortwijk, 1995; Timmermans et. al., 1992).

Each of the chosen stated and revealed preference models — discrete-choice hedonic analysis

---

[3] The main drawback of discrete-choice hedonic analysis is the need to impose a good deal of structure on the utility function (Cropper et. al., 1993).

[4] Louviere (1988) provides a nice review of this type of analysis and its application.

and choice-based conjoint analysis — has its advantages and disadvantages. The common criticism of any stated preference method is the hypothetical nature of the questions and people's choices (Mitchell and Carson, 1989). The main strength of any revealed preference method is that it is based on observed behavior. However, the revealed method of hedonic analysis suffers from several weaknesses. First, hedonic analysis depends critically on the control of all important structural, neighborhood, and environmental factors behind location choices (Freeman, 1993). To cope with this dependence, previous studies incorporate numerous explanatory variables, yet may still omit important variables. Second, hedonic analysis suffers from collinearity between explanatory variables, especially when many are included (Freeman, 1993); this aspect precludes the isolation of factors affecting housing choice. Moreover, collinearity generates coefficients with wrong signs or implausible magnitudes (Greene, 1993). Third, hedonic analysis of actual housing purchases is unable to capture effectively the influence of uncommon attributes or unusual levels of attributes. These first three weaknesses apply to both types of hedonic analysis.

Additional weaknesses apply particularly to discrete-choice hedonic analysis. The most relevant to this study involves the definitions of alternative housing locations available to the individual household — the feasible choice set. Given limited information on households' search strategies, any analysis of housing purchases requires the researcher to specify arbitrarily the feasible choice sets of housing alternatives that were considered by individual households. Moreover, the size of the specified feasible choices set may be computationally intractable, forcing the analysis to reduce dimensionality through information-depleting means.[5]

[5] Anas and Chu (1984) provide an excellent overview of the error sources and estimation biases associated with discrete choice models, with a focus on their application to housing and travel mode choices.

Choice-based conjoint analysis avoids each of these weaknesses. First, the choice sets of conjoint analysis specify the attributes associated with each housing alternative; this specification clearly identifies the parameters to consider when choosing a house. Second, the statistical design of choice-based conjoint analysis avoids collinearity by generating orthogonal attribute data; i.e., the level of one attribute is held fixed, while the level of another attribute changes. Third, the survey design of conjoint analysis generates an adequate number of observations for all attributes and attribute values, including the uncommon ones. Fourth, conjoint analysis prespecifies the alternatives within each choice set faced by households.

By combining the stated and revealed preference methods, the joint model enhances the strengths and diminishes the drawbacks of each individual method. This combined approach yields three benefits. First, the statistical design of choice-based conjoint analysis generates orthogonal attribute data (e.g., hold constant the number of bedrooms, while increasing the number of bedrooms). The addition of stated preference data reduces the collinearity that most likely exists in the revealed preference data on housing choices. Consequently, estimation is able to identify attribute effects that would be obscured by collinearity. Second, the stated preference questions generate additional observations for attributes or attribute values that are uncommon within the revealed data. Third, inclusion of revealed preference data ensures that estimation is based on observed behavior to some degree.

Fortunately, these two methods can appropriately be combined since they reflect the same process of selecting a housing location based on attributes. As constructed, both models are discrete choice models. Therefore, discrete choice random utility theory and multinomial logit estimation techniques apply to both models and generate comparable estimates (Cropper et. al., 1993).

3. Theoretical Framework

This paper employs random utility theory to model individuals' choice among housing location alternatives for both the hedonic analysis — observed choice from an actual choice set — and the conjoint analysis — induced choice from a hypothetical choice set. In both analyses, the individual (indexed by *n*) chooses the housing location that yields the highest utility of all locations in the feasible set K_n.

In the random utility framework, overall utility, U_{in}, is the sum of a deterministic component, V_{in}, and a random component, e_{in}:

$$U_{in} = V_{in} + e_{in},$$

where *i* identifies the location. I model the deterministic component as an indirect utility function conditional on the following arguments:

Z_i = vector of observed housing location attributes,

C_n = vector of observed individual characteristics,

y_n = income of individual *n*,

P_i = price of location *i*, and

β = parameter vector to be estimated.

In other words, $V_{in} = V_{in} (y_n - P_i , Z_i , C_n ; \beta)$. The random component (or error term) may reflect (1) unobserved attributes of the individual or housing location or (2) deviations in individual *n*'s preference vector β_n from the mean preference vector β; i.e., unobserved heterogeneity in preferences (Cropper et. al., 1993). If the error terms are identically and independently distributed (IID) Type I Extreme value with scale parameter μ,

μ = scale parameter,

the probability that individual *n* chooses location *i* rather than location *j* is of the logit form:

$\pi_n(i)$ = probability that individual *n* chooses location *i* rather than location *j*,

$= P(V_{in} + e_{in} \geq V_{jn} + e_{jn} : \forall j \in K_n)$,

$= \exp(\mu V_{in}) / \sum_{j \in K} \exp(\mu V_{jn})$.

This equation represents a well-behaved probability bounded between zero and one (Quigley, 1985). If the deterministic utility component of the utility function is linear in its parameters,

$$V_{in} = \beta_0 + \beta_Z Z_i + \beta_C C_n + \beta_y (y_n - P_i),$$

where $\beta = \{\beta_0, \beta_Z, \beta_C, \beta_y\}$, then estimated parameters are unique up to the scale factor μ (McFadden, 1978). Empirical analysis generally assumes this factor equals one. Since this study examines two separate data sets, it is able to estimate jointly the relative scale factor (i.e., ratio of the two scale factors) for one data set along with the model parameters for the joint data (Adamowicz et. al., 1994).

This structure assumes that the odds of choosing housing unit *i* relative to unit *j* are independent of the attributes of all other housing alternatives — independence of irrelevance alternatives (IIA). While this assumption may be inappropriate in many situations involving the choice of housing locations (Quigley, 1985), models that include many socioeconomic attributes in an appropriate fashion may generate reasonable estimates since the deterministic component of the utility function should account for population heterogeneities (Ben-Akiva and Lerman, 1985).

A further complication involves selection of the feasible set of housing alternatives. In the conjoint analysis, the feasible set consists of the three constructed housing alternatives. However, in the hedonic analysis of actual housing choices, consumers select one specific housing location from a large number of alternative locations actually available on the market, K_n. In order to keep the analysis tractable, one must reduce the size of the choice sets. By selecting a subset of alternatives, noted *d*, and observing each household's selection among locations within this subset, regression analysis obtains consistent estimates of the correct choice model (Quigley, 1985). Let $f(d|i)$ represent the sampling rule for obtaining subset *d*, conditional upon the observed selection of housing unit *i*. McFadden (1978) shows that if the sampling rule has the "uniform conditioning property," maximization of the likelihood function based on a sample of observations on choice *i* from the subset *d* yields the <u>same</u> consistent parameter estimates obtained by maximizing the likelihood function based on observations of choice *i* from the set of <u>all</u> possible alternatives, K_n. The following sampling rule has this helpful property: choose *d* by including the chosen alternative and selecting at random ω rejected alternatives in the feasible set (Quigley, 1985); put differently,

$$f(d|i) = \omega / (N_n - 1),$$

where N_n indicates the number of elements in the feasible set K_n.

For the empirical analysis of the Fairfield housing market, the feasible set consists of all locations sold in the town during the same month and year. It seems reasonable to assume that any household could feasibly live anywhere in the study area given its small size (Nechyba and Strauss, 1998). Also, the number of randomly drawn alternatives, ω, equals three in the empirical analysis. Parsons and Kealy (1992) show that even a limited number of alternatives, as small as three, is appropriate for randomly drawn opportunity sets in a random utility model.

4. Analytical Approach

Given this theoretical framework, the following section depicts two separate approaches to analyzing residential location choices: discrete-choice hedonic analysis of revealed data and choice-based conjoint analysis of stated data. Section 5 further develops these two approaches and depicts a third analytical approach: joint analysis of combined data.

4.1. Discrete-Choice Hedonic Analysis

4.1.1. Research Framework

The discrete choice hedonic model views the individual as choosing the housing location that gives him/her the highest utility out of all the housing locations available in a universal choice set (Cropper et. al., 1993). In this view, utility is a direct function of the housing location attributes. The previous literature on hedonic analysis includes many attributes or factors influencing housing location choices (Cropper et. al., 1988; Palmquist, 1992). These factors divide into three main categories: structural, neighborhood, and environmental. This analysis includes the following prominent structural features:

(1) style,

(2) number of bedrooms,

(3) number of bathrooms,

(4) interior space,

(5) lot size, and

(6) age of structure.

This analysis includes two neighborhood features. First, it includes indicator variables for two of the most prominent neighborhoods in Fairfield — "the beach" and Greenfield Hills — using census tract boundaries. [Frech and Lafferty (1984) also use census tract boundaries to distinguish neighborhoods.][6] Second, it controls for flooding frequency, which is quite relevant for Fairfield

[6] The "beach" includes the two census tracts with waterfront property on Long Island Sound. Greenfield Hills is a historically prestigious area of Fairfield with relatively spacious houses on rather large estates of land abundant with trees and horse pastures.

given that much of the town is built on former coastal wetland (Steadman, 1996). Otherwise, this analysis ignores most neighborhood features because the study site involves only a single small town (population approximately 40,000) that is relatively homogenous in terms of the neighborhood features employed in previous research: percent professional, median income of census tract, percent of houses owner-occupied, percent white, and median age of census tract.

Most environmental attributes employed in previous research, such as air quality (Graves et. al., 1988), vary only minimally due to the small study area. Nevertheless, the town of Fairfield generates a strong variation in the environmental amenity or natural feature associated with (i.e., immediately adjacent to) a given housing location. In this analysis, the natural feature takes one of the following eight values:

Water-Based Features:

(1) Long Island Sound,

(2) saltwater marsh,

(3) freshwater marsh,

(4) river or stream,

(5) lake or pond,

Land-Based Features:

(6) forest or woods,

(7) open field or park,

No Feature:

(8) backyard lawn.

Actually, the category of backyard lawn establishes the <u>absence</u> of a natural feature.

In addition to attributes associated with the housing location itself, this study also incorporates information on the characteristics of the home buyer: marital status, presence of dependent children living at home, size of household, and annual household income. This information helps to explain housing choices since it captures potential heterogeneity in individuals' housing demands and abilities to pay.

Since these factors may not sufficiently control for variation in housing locations, this analysis attempts to incorporate the "un-measured quality" associated with each housing location using hedonic price analysis (Ellickson, 1977). Using the same data examined for the discrete-choice hedonic analysis, this approach regresses the price of each housing location on the same set of structural, neighborhood, and environmental attributes included in the discrete-choice hedonic analysis.[7] The price residual calculated for each housing location captures "un-measured quality;" i.e., it represents an index of those aspects of housing quality not captured by the vector of attributes.

4.1.2. Data Collection Methods

Data on actual housing choices, their associated attributes, and characteristics of buyers are taken from several sources. The Town of Fairfield Tax Assessor records all housing purchases transacted in the town of Fairfield. A computer database supplied by this office provides the following information on housing purchases:

(1) style,

(2) number of bedrooms,

(3) number of bathrooms,

---

[7] This hedonic price approach technically regresses the log value of house price on the explanatory variables. In this way, the residual is not a linear combination of the explanatory variables included in the discrete-choice hedonic analysis.

(4) interior space (in square feet),

(5) lot size (in acres),

(6) age of structure,

(7) date most recently sold,

(8) location (i.e., street address),

(9) name of new owner, and

(10) purchase price.

From this database, I collected data on the prominent structural features employed in previous hedonic studies. The database contains numerous types of houses: single-family residences, multi-family residences, condominiums, etc. To avoid the need of differentiating housing markets among these different types, this paper examines only privately-owned residential single-family dwellings.

Given the street address, I was able to collect data on the natural feature associated with each housing location. The Natural Resources Center of the Connecticut Department of Environmental Protection provides data on land use and land cover for the entire town of Fairfield. The Town of Fairfield Tax Assessor provides data on street addresses for each land parcel in the town of Fairfield. By overlaying these data and examining other topographical maps, I identified the most likely natural feature associated with each housing location. Then I verified or modified the natural feature through on-site inspection at <u>each and every</u> housing location.

Information on street address also allowed the identification of flooding frequency for each particular housing location. The Town of Fairfield Planning and Zoning Commission provides information on flooding classifications for the entire town of Fairfield. By overlaying these data with data on street addresses, I classified each housing location according to three categories:

(1) subject to the 100-year flood,

(2) subject to the 500-year flood, and

(3) subject to minimal flooding.

Information on individual homeowners' characteristics is elicited through mail surveys. This collection method is described in Section 4.2.2.

4.2. Choice-Based Conjoint Analysis

4.2.1. Research Framework

Choice-based conjoint analysis attempts to mimic the discrete choice hedonic analysis. Rather than observing people's choice from an actual set of housing alternatives, choice-based conjoint analysis asks people to choose from a hypothetical set of housing alternatives, which vary according to the associated attributes. The attributes used to describe each alternative reflect the actual characteristics of housing locations in the study area; Table 1 displays these attributes. (Conjoint analysis excludes the "neighborhood" attribute because it is difficult to present within a survey context.) Moreover, the analysis bases the values for each attribute on the actual ranges of values for housing locations in the study area. The statistical design process used to generate the choice sets requires discrete attribute levels. For some attributes, the variables are inherently discrete (e.g., house style). In these cases, I selected the most frequent categories found in the revealed preference data in order to span a reasonable portion of the market. For other attributes, the variables are inherently continuous (e.g., lot size). In these cases, I selected "rounded" values near the first-quartile, median, and third-quartile levels of the revealed preference data, as appropriate. For example, the first-quartile value for purchase price is $ 182,000; the value included in the choice set design process is $ 200,000. Table 1 displays the values included for each attribute.

In the conjoint survey, each choice set includes three housing alternatives. These alternatives are based on the natural feature associated with the housing location: water-based feature, land-based feature, and no natural feature. (Backyard lawns are viewed as a feature that is not truly "natural.") Figure 1 shows an example taken from this portion of the conjoint survey.[8] The survey need not divide the choice set into categories; alternatively, the survey could identify the alternatives merely by number (e.g., House # 1, House # 2, etc.). The chosen design reduces the number of choice sets sufficient to estimate consumer preferences, as explained in the next paragraph. Of course, division of the housing alternatives could be accomplished using other housing attributes besides natural feature, such as style. This survey is designed to serve two research projects; the other project attempts to measure the aesthetic benefits generated by each type of natural feature. The chosen division facilitates this other project.

The set of attributes and levels displayed in Table 1 can be seen as establishing the space to be spanned in the choice experiment (Adamowicz et. al., 1994). Given that one views each attribute as discrete, there exist (2^2 x 3^3 x 4^2 x 5^2) possible water-based alternatives, (2^3 x 3^3 x 4^2 x 5) possible land-based alternatives, and (2^2 x 3^3 x 4^2 x 5) possible no-feature alternatives. Consequently, one can view the issue of choice set construction as sampling from the space of possible triplets of water-based, land-based, and no-feature alternatives (Adamowicz et. al., 1994). Assuming that the choice process can be depicted by McFadden's (1975) "Mother" logit model, the design strategy described

[8] Timmermans and van Noortwijk (1995), one of the two previous applications of choice-based conjoint analysis to housing choices, include two alternatives and a third "no purchase" option. Without this third option, the construction of housing alternatives assumes the conditional logit model applies, in other words, one of the choices is acceptable to each respondent. The inclusion of a "no purchase" option is not appropriate for matching the stated data with the available revealed data on housing purchases since a household always buys a home. Moreover, the greater is the number of alternatives, the more realistic is the choice set.

here is consistent with a subset form of the more general Mother logit form (Adamowicz et. al., 1994; Louviere and Woodworth, 1983; Louviere and Hensher, 1983). In this design strategy, I first treat the attributes of water-based, land-based, and no-feature alternatives as a collective factorial — (2^2 x 3^3 x 4^2 x 5^2) x (2^3 x 3^3 x 4^2 x 5) x (2^2 x 3^3 x 4^2 x 5). Then I use an orthogonal main effects design that varies simultaneously all the attribute levels; i.e., the attributes of the choice alternatives are orthogonal within and between alternatives.[9] This design permits the consistent estimation of the strictly additive variance components of the Mother logit model, given that all interactions are zero; however, the design does not generate optimally efficient parameter estimates (Adamowicz et. al., 1994). Still, it produces relatively efficient estimates (Bunch et. al., 1992).

4.2.2. Data Collection Methods

The main effects design demands 81 choice sets, derived from the (2^7 x 3^9 x 4^6 x 5^4) full factorial of potential attribute level combinations. Few individuals would be willing to respond to all 81 choice sets in a mail survey. Therefore, I randomly divided the 81 choice sets into 9 groups of 9 choice sets each.[10] I placed each group of nine choice sets into a similar survey format. In other words, I generated nine versions of the same survey format, each containing nine choice sets.

---

[9] Adamowicz et. al. (1994) notes logit models are "difference-in-utility" models, that is, parameters are defined by differences in attribute levels. The statistical design employed in this study orthogonalizes the absolute attribute levels but not the differences. (Nevertheless, the logit model applies.) Inclusion of a constant reference alternative to each choice set preserves the orthogonality, even in differences, by providing a constant point for calculation. However, no constant reference is appropriate for matching the stated data with the revealed data on actual purchases since no one alternative was available to all buyers.

[10] Rather than randomly dividing the 81 choice sets, I could have blocked them into 9 groups by using an additional four-level column as a factor in the main effects design. This blocking procedure guarantees that every level of every attribute is represented in each group. Computer limitations at a critical juncture unfortunately precluded this better procedure.

The complete survey consists of four parts.[11] Part one introduces and briefly explains the research project. Part two visually depicts the eight natural features using digitally scanned black-and-white photographs. (See Figure 2.) By visually depicting rather than verbally describing the natural features, this study reduces the perceptional variation across respondents. In other words, all respondents have the same visual image for a given natural feature. Part three collects information on contingent behavior by asking the respondents to imagine that they must leave their current home and choose among three possible new housing locations. (See Figure 1.) Part four requests information on the respondents' characteristics.

This research project mailed 499 mail surveys (evenly distributed across the nine survey versions) to homeowners in the town of Fairfield, CT, in late 1996. The names and addresses of potential respondents were taken from the house purchase database provided by the Town of Fairfield Tax Assessor. The database includes all sales contracted between January 1994 and August 1996, inclusively. For this period, the sample of privately-owned residential single-family dwellings includes 1,501 houses. Then I applied a stratified random sample selection process, within which I oversampled houses located close to Fairfield's coastal marshes by including all such houses (130 houses) in the final mailing sample. This oversampling attempts to include housing locations with natural features that represent a small proportion of the housing market. Then I randomly selected 369 houses not located adjacent to a coastal marsh from the possible 1,371 non-marsh-adjacent houses. Of the 499 people contacted, 105 returned completed surveys, for a response rate of 21 %.

4.3. Data

This sub-section examines the data collected on households and their actual and hypothetical

[11] A copy of the survey is available from the author upon request.

housing purchases and selections. Table 2 summarizes the characteristics of the households included in the sample. The modal household has married parents and children in a family of two people (obviously contradictory) earning between $ 100,000 and $ 200,000 in annual income. Keep in mind that the town of Fairfield resides in one of the wealthiest counties in the US.

Table 3 summarizes the data on housing alternatives considered by households when actually purchasing a home and when posed with hypothetical choice sets. The first two columns report the attributes of the sampled houses within the revealed data model; these columns report chosen and rejected houses, respectively. On average, chosen housing locations cost $ 304,000, were 44 years old, contained 3.3 bedrooms and 2.3 bathrooms, provided nearly 2,000 square feet of interior space, and sat on 0.7 acres of land. The modal chosen house was a Colonial with no natural feature (i.e., backyard lawn) located in a less prominent neighborhood facing minimal flooding risk. The housing locations rejected by the sampled households, on average, are cheaper, older, and tied to more land, yet comparable in terms of interior space, bathrooms, and bedrooms. Also, rejected houses are more likely to lack a natural feature, more prone to flooding, and less likely to exist in a prominent neighborhood. Without controlling for multiple variation in the attributes, these data indicate that households' choices are consistent with our expectations regarding natural beauty, flooding risk, and neighborhood prominence, yet inconsistent regarding cost and lot size. The age of a house may connote more depreciation and/or less panache; thus, the effect of age on household choice remains ambiguous.

The third and fourth columns of Table 3 report attributes of hypothetical houses within the stated data model; these columns report chosen and rejected houses, respectively. On average, chosen hypothetical houses cost $ 282,000, were 34 years old, contained 1.4 bathrooms and 3.3

bedrooms, provided about 1,900 square feet of interior space, and sat on 0.4 acres of land. The modal chosen hypothetical house was a Colonial with no natural feature in a region of minimal flooding. The hypothetical houses rejected by household respondents, on average, are more expensive, older, and less spacious, yet comparable in terms of lot size, bathrooms, and bedrooms. Also, rejected hypothetical houses are more likely to lack a natural feature and more likely to be styled as a Cape Cod or Ranch, yet comparable in terms of flooding risk. Without controlling for multiple variation in the attributes, these data indicate that households' choices are consistent with our expectations regarding cost, interior space, and natural beauty. The effect of housing style is certainly ambiguous.

Comparison of the revealed and stated data reveals an overall comparability between the two data sets. Nevertheless, a few characteristics differ. Relative to actual houses, hypothetical houses are newer, sit on smaller lots, possess fewer bedrooms, and more likely to possess natural features, especially water-based features. This last difference is the most prominent; this great disparity explains why the sampling process of actual housing locations is stratified according to natural feature. These differences may lead the estimation procedures to reject the hypothesis of similar decision processes guiding actual and hypothetical housing selections, in other words, reject the notion of identical parameter estimates. Section 5 revisits this point.

Further exploration of the revealed data confirms the expectation of multicollinearity inherent between housing attributes. As shown in Table 4, the expected culprits are strongly and significantly correlated. Price is positively correlated with lot size, interior space, number of bedrooms, number of bathrooms, and neighborhood prominence. Age is positively correlated with the number of bathrooms, interior space, and neighborhood prominence. Lot size and interior space are correlated.

Bedrooms and bathrooms are correlated.[12] Consequently, the coefficients of these individual variables may prove to be insignificant, take unexpected signs, and/or take implausible magnitudes (Greene, 1993). This high potential for multicollinearity, in combination with the inclusion of individual attributes and the *ceteris paribus* condition of regression analysis, makes the interpretation of individual coefficients difficult. In particular, these conditions may change the sign of a coefficient from the expected effect of an attribute if it were considered in isolation (Frech and Lafferty, 1984). For example, additional bedrooms in a house of fixed interior space may generate little or negative value. As stated above, choice-based conjoint analysis avoids this multicollinearity.

5. Econometric Analysis

This section analyzes the collected data on actual and hypothetical housing choices and attempts to identify the factors driving these choices. In particular, it looks to combine the revealed and stated preference data in an effort to improve the identification process and increase our understanding of residential location choices. If regression analysis finds that the two data sets capture similar decision processes (i.e., the data sets are compatible), then the combination of stated and revealed data should diminish the weaknesses, while enhancing the strengths, of each individual analytical method. If regression analysis indicates that the two data sets are incompatible, further analysis can isolate the factors causing the two decision processes to diverge. Then future research can explore more deeply these incompatible parameters.

5.1. Structure

Given the assumptions of the random utility framework structured in Section 3, this paper

---

[12] Only the more interesting correlations are noted in the text and reported in Table 4. Complete tabulation of Pearson correlation coefficients and their statistical significance is available from the author upon request.

applies the multinomial logit model and estimates the parameter vector β associated with deterministic utility using full-information maximum likelihood techniques (Cropper et. al., 1993). Due to the stratified random sampling design, I weight the observations according to their different likelihoods of entering the estimation.[13] When estimating the stated data, the replications of choices from individual respondents are assumed independent, a common practice when examining stated choice data (Adamowicz et. al., 1994; Adamowicz et. al., 1997; Swait and Adamowicz, 1996).

Estimation demands a few further details. First, I employ 1,0 dummies for two of the three broad natural feature categories: water-based and land-based (no-feature is the benchmark category). These dummy variables represent alternative-specific constants in the conjoint model but not the hedonic model. Second, I employ effect codes rather than 1,0 dummies to distinguish all other attributes with multiple levels (e.g., house style), as is conventional in conjoint analysis.[14] This specification improves the interpretation of coefficients involving interactions and does not confound the estimation of the alternative-specific constants. [See Adamowicz et. al. (1994, pg. 280-281) for the full rationale behind this specification.] Third, I interact the explanatory parameters regarding household characteristics with the price of each housing alternative. Otherwise, these explanatory parameters do not vary within each household's choice set. In addition, this interaction allows the effect of price to vary across households with presumably differing abilities to pay.[15] Fourth, effects

[13] Estimation of this weighted exogenous sample maximum likelihood function generates consistent estimates; however, they are not asymptotically efficient (Ben-Akiva and Lerman, 1985).

[14] Each level of the attribute except the base level is represented by a column. Each column contains a "1" for the level represented by the column and a "-1" for the base level. The interpretation of these parameters is that the base level takes the utility level of the negative of the sum of the estimated coefficients and each other level takes the utility associated with the coefficient (Adamowicz et. al, 1994).

[15] Other types of interaction serve the purposes of incorporating household characteristics and addressing interesting aspects of housing choice. For example, the effect of interior space may depend on

codes capturing different years prove to be statistically insignificant for the hedonic analysis and do not apply to the conjoint analysis.

5.2. Estimation

To estimate the parameter vector of deterministic utility, I employ three separate sets of data: only revealed preference data, only stated preference data, and combined data.

5.2.1. Separate Estimation of Revealed and Stated Preference Data

Revealed Preference Data

This sub-section estimates household utility using each type of data separately. First, it estimates household utility using only revealed data on actual house purchases. Estimation results are shown in Table 5.[16] The likelihood ratio statistic indicates a highly significant collection of coefficients, yet McFadden's ρ^2 indicates only a reasonable fit of the data (ρ^2=0.33). Several of the parameters, though not most (11 of 25), have significant effects on household utility. Most of the parameters take the expected sign. Where relevant, the relative magnitudes are generally appropriate.

Estimation generates the following particular results. Households are more likely to purchase spacious houses, Colonial-style houses (relative to Cape Cods) or house styles other than Cape Cod, houses in the Greenfield Hills neighborhood, or houses with more "un-measured quality" — as captured by the hedonic price residual. (Given the large magnitude and strong significance of this last coefficient, the need for this parameter is quite apparent. The other attributes representing measurable characteristics seem to capture rather incompletely housing "quality.") Households are

---

household size or the presence of children. However, these specifications of interactions generate less satisfying results, which are available from the author upon request.

[16] Estimation without weighting generates substantially different coefficient estimates for only four variables and statistical significance for only five variables.

less likely to buy a more expensive house. And as one would expect, the price of a house has less effect on a household with medium income than a one with low income and even less effect on a high-income household. Finally, consider the effect of natural features. Households are more likely to buy houses with water-based natural features than one with no natural feature. Within the broad category of water-based features, households are more likely to buy houses adjacent to rivers/streams and saltwater marshes and less likely to buy ones adjacent to freshwater marshes.

As predicted, a few important parameters have statistically insignificant coefficients, some parameters take inappropriate relative magnitudes, while several others take coefficient signs opposite from expectations (none being significant). First, consider statistical insignificance. Although their coefficients take the expected sign, the number of bathrooms, age,[17] exposure to 500-year floods, and land-based features do not significantly affect housing choices. In addition, houses adjacent to Long Island Sound are not significantly more attractive than water features as a group. Second, consider inappropriate relative magnitudes. Households are less likely to buy houses adjacent to Long Island Sound than those adjacent to either rivers or saltwater marshes. Third, consider unexpected signs. Households are less likely to buy houses with more bedrooms or larger lots, yet more likely to buy houses exposed to 100-year floods. These surprising and odd results can be explained by the strong correlations between these attributes and other important attributes.

---

[17] As noted in Section 4.3, age may generate two countervailing influences on deterministic utility. An increase in age most likely degrades the structural quality of a house, yet improves its “character”. One means of capturing these two effects is to incorporate both age and age-squared into the specification of deterministic utility. Estimation of this specification found the effect of age-squared insignificant; therefore, the final specification excludes this variable.

Stated Preference Data

In an attempt to improve the analysis on all three counts, this sub-section next estimates household utility using only stated data on hypothetical house purchases. Estimation results are shown in Table 6.[18] Relative to the regression of revealed data, the likelihood ratio statistic indicates a more significant collection of coefficients, yet McFadden's ρ^2 indicates a worse fit of the data (ρ^2=0.20). Relative to estimation of the revealed data, many more parameters are significant (15 of 21). Unlike the revealed data, the relative magnitudes are completely appropriate, where relevant, and none of the coefficients take unexpected signs.[19]

Estimation reveals the following particular results. Households are more likely to select houses with more bathrooms, more interior space, or larger lots. On all three counts, these results improve upon the revealed data: significant correct sign, greater significance, and significant correct sign, respectively. Also, households are more likely to select a Colonial-style house (relative to a Cape Cod), yet less likely to select a Ranch-style house or an older house. As with the revealed data, the price of a house has less effect on medium-income households than a low-income household and even less effect on a high-income household. Oddly, price alone has no significant effect on housing choices, unlike the revealed data. However, under numerous other specifications, this effect is extremely significant. In particular, this effect remains significant as long as the data are not weighted according to the stratification proportions or interactions with marital status and household size are

[18] Estimation without weighting generates substantially different coefficient values for only two variables and statistical significance for only four variables. Complete results are available from the author.

[19] The estimated interaction between price and marital status may be questionable depending on one's expectation. Estimation indicates that married households respond more negatively to the price of a house. On one hand, marriage may reduce the resources available to a household given a fixed level of income and household size. On other hand, this analysis only crudely controls for income within rather large brackets and does not control for working adults. Therefore, marriage generally should increase household resources and reduce the negative effect of price on housing selections.

excluded. In the latter case, the interactions seem to absorb the explanatory power associated with the housing price.

Finally, consider the effects of natural features. Households are more likely to select houses located adjacent to water-based or land-based features than houses lacking a natural feature. Within the broad category of water-based features, households are more likely to select houses near Long Island Sound, rivers/streams, and lakes/ponds, yet less likely to select houses near freshwater and saltwater marshes. (The latter effect is insignificant.) Within the broad category of land-based features, households prefer forests over open fields. Relative to the revealed data, these results show an improvement in the identification of land-based features, forest versus open field, and Long Island Sound. Besides this increase in statistical significance, these estimated effects of natural features differ from revealed data results in only one respect: saltwater marshes positively (and significantly) influence actual housing purchases, yet they negatively (and insignificantly) influence hypothetical housing selections.

5.2.2. Joint Estimation of Revealed and Stated Data

All Coefficients Constrained to Be Equal Across the Two Types of Data

Application of the multinomial logit / maximum likelihood techniques to the first two sets of revealed and stated data is straightforward. Application to the combined data demands further comment since it involves a joint estimation procedure, which permits estimation of the relative scale factor, μ, for the two data series. Swait and Louviere (1993) describe the appropriate steps to joint estimation. First, separately estimate the revealed model and the stated model. The log-likelihood values for these models are L_r and L_s , respectively. Second, concatenate the two data sets and estimate the joint model. Revealed and stated data are assumed independent, a common practice when combining these types of data (Adamowicz et. al., 1994; Adamowicz et. al., 1997; Swait and Adamowicz, 1996). The log-likelihood value for this model is L_n. Third, concatenate the two data sets but rescale the stated data relative to the revealed data (or vice versa) by conducting a grid search:

(a) multiply the stated data matrix by a constant, beginning at one end of the search range;

(b) estimate the joint model and its log likelihood, denoted L_c;

(c) repeat by incrementing the constant; and

(d) stop at the constant value that maximizes the likelihood value.

This procedure generates the optimal rescaling constant that maximizes the fit of the stated and revealed parameters given the conditional logit model (Adamowicz et. al., 1994). (In this particular application, the optimal rescaling factor equals 49.7.) Fourth, use these log-likelihood values to examine whether the preference structures are similar between the two data sets by testing the hypothesis of equal parameters, after adjusting for the relative scale effect. In other words, use the following likelihood ratio test of the difference between parameters: $\lambda = -2[L_c - (L_r+L_s)]$. Failure to reject this χ^2 test would provide sufficient evidence that the stated and revealed data contain similar preference structures. In this analysis, the calculated χ^2 test statistic, λ, for housing location choices equals 91.446. Given 31 degrees of freedom,[20] this test statistic significantly rejects the hypothesis of equal parameters at the 1 % confidence level. In other words, the revealed and stated data are not compatible.

Should this finding surprise us? Some previous research on combining revealed and stated

[20] The degrees of freedom equal the number of parameters in the revealed data model plus the number of parameters in the stated data model minus the number of parameters in the joint model plus one additional degree for the relative scale factor (Swait and Louviere, 1993).

data reports similar findings of incompatibility (Swait and Adamowicz, 1996; Adamowicz et. al., 1997). Several other analyses support the notion of compatibility (Huang et. al., 1997; Adamowicz et. al., 1994; Louviere, 1996), but these analyses examine exclusively recreational and transportation choices. No previous analysis attempts to combine stated and revealed data on housing choices, as noted in the introduction. For several reasons, the finding of incompatibility may not be surprising for the analysis of housing choices. First, the omission of relevant variables from the utility function may erroneously lead to the conclusion of different parameter vectors in two data sets (Swait and Louviere, 1993). Certainly, this potential is quite high in the case of residential choices. Second, the rejection of equal parameters may stem from the probable difference in the entropy of choice sets between the revealed and stated data, where entropy is measured by the degree of closeness among all alternatives within a given choice set (Swait and Adamowicz, 1996). The less obvious is the dominance of one alternative over another, the greater is the entropy. The choice sets of actual housing alternatives most likely involve greater entropy since they display more variation in the attribute levels, especially the attributes of interior space and lot size (as shown in Table 3). Put differently, alternatives become more similar when trade-offs become more difficult to evaluate, which is the case under greater variation. Third, dramatically different distributions of attribute levels between the two data sets may disrupt the testing of equal parameters. As shown in Table 3 and noted in Section 4.3, the distributions of certain attributes dramatically differ between the two data sets. Fourth, a large difference in variance magnitudes between the two data sets may disrupt the testing of equal parameters. The optimal rescaling factor equals the ratio of revealed data variance to stated data variance; a factor value of 47.9 clearly indicates that the revealed data involves more variance than the stated data since this factor. The variance in revealed data is expected to be higher since real market choices are probably subject to many more random influences (Louviere, 1996).[21] Finally, the decision processes surrounding actual and hypothetical housing choices may truly differ. The hypothetical context of the conjoint survey may simply not replicate the reality of buying a home.[22] This final point is revisited.

Only Compatible Coefficients Constrained to Be Equal Across the Two Types of Data

The analysis so far does not distinguish exactly where the two decision processes diverge. Thus, a question still remains: Which factors or attributes make the two data sets incompatible? The effects of certain attributes may be comparable between the two data sets, while the effects of other attributes may be completely different. In order to separate compatible and incompatible variables, I allow certain subsets of the coefficients to vary between the two data sets when estimating the joint model. In other words, the two data sets are pooled, yet certain coefficients are not restricted to be equal across the two data sets. The strategy is to identify the largest subset of variables constrained to be equal across the two data sets which does not reject the hypothesis of parameter estimates. To illustrate this search, define the following notation:

W_x = set of all variable collections consisting of "x" variables restricted to be equal across the two data sets, where $w_x \in W_x$; and

L_x = log-likelihood value associated with variable collection w_x.

---

[21] The magnitude of the variance ratio is however extremely large. Further analysis shows that the variable of house price drives this high value. Once the effect of price is allowed to vary between the two data sets, the optimal rescaling factor drops to a very reasonable value of 1.31. This value still indicates that variance in the revealed data is greater.

[22] Given the limited number of attributes, the survey may cause respondents to focus on attributes that they do not consider important when actually purchasing a home, such as flooding risk. If true, the effects of such attributes may differ between the revealed and stated data.

To implement the search, first determine the collection of variables that minimizes the log-likelihood value for each subset size:

$w_x^* = \text{argmin } \{L_x: w_x^* \in W_x\}$,

L_x^* = log-likelihood value associated with variable collection w_x^*, and

λ_x^* = chi-square test statistic associated with L_x^*.

Then the largest collection of variables not rejecting the hypothesis of equal parameters is x*:

$$x^* = \text{argmax } \{x: \lambda_x^* \leq \chi^2_{c,m}\},$$

where $\chi^2_{c,m}$ denotes the critical value of the chi-square distribution at the c-% confidence level for "m" degrees of freedom.

This method finds that 12 particular variables represent the largest collection of compatible variables. Eight variables remain unrestricted in their effects between the two data sets. Six variables are not common to both sets so they are regarded as neither compatible nor incompatible. Table 7 lists the variables in each category. According to this method, the following attributes seem to cause the two housing decision processes to diverge: price, its interaction with marital status and household size, saltwater marsh, the housing styles of Colonial and Ranch, lot size, and the risk of 100-year flooding. A likelihood ratio test confirms that these variables differ in their effects between the two data sets; in other words, the hypothesis of equal parameters is rejected at the 1.0 % significance level with a χ^2 statistic of 64.83. The remaining variables generate similar effects between the two data sets, given an adjustment in scaling adjustment. The optimal rescaling factor for the joint model generating these results equals 0.85.

This scale factor of 0.85 indicates that variance in the stated data exceeds the variance in the revealed data once certain parameters are not restricted to be equal across the two data sets.

Moreover, the difference in scale factor levels, 0.85 versus the initial value of 47.9, indicates that the factors causing incompatibility between the two data sets are associated with greater variance in the revealed data.

Given that the hypothesis of equal parameters is not rejected, one can test whether rescaling one of the two data series is justified by calculating a likelihood ratio test of equal scale factors, conditional on the assumption of equal parameters: $\delta = -2[L_n - L_c]$. Rejection of the null hypothesis of equal scale factors justifies the rescaling. The χ^2 test statistic for δ equals 0.256. Given one degree of freedom, the hypothesis of equal scale factors is not rejected. This result indicates that the hypothetical situation is quite similar to the real market (Louviere, 1996), given the differences in the incompatible parameters.

Estimation of this specification for combining revealed and stated data generates the results shown in Table 8. The likelihood ratio statistic indicates a collection of coefficients more significant than either separate data set. Unfortunately, McFadden's ρ^2 indicates only a slightly better fit of the data (ρ^2=0.21) than the stated data alone and a worse fit than the revealed data alone. This disappointing result may stem from the presence of eight incompatible parameters. Examination of the individual coefficients reveals encouraging results. As predicted, combining revealed and stated data improves identification of influential housing attributes and household characteristics. Estimates of the compatible parameters identify the significant and appropriately signed effects of broad and individual categories of natural features, bathrooms, interior space, age, and household income. Estimation of the revealed data alone fails to identify appropriately the effects of certain natural feature categories, bathrooms, and age. Of course, the mere inclusion of extra observations from the relatively larger stated data set may drive this result. However, this point carries so much weight since the effects of the parameters are still found to be similar between the two data sets. Estimates of the incompatible parameters reveal more interesting but less obvious improvements in identification. The results indicate that the inclusion of stated data also allows the identification of important parameters even within the revealed data itself. Estimation of only revealed data fails to identify the effects of Ranch-style, marital status, and household size. However, addition of the stated data permits identification of these important effects.

6. Summary

This paper combines revealed and stated preference methods and data to examine housing choices. This combination enhances the strengths, while diminishing the drawbacks, of each individual method, thus improving the identification of influential parameters, especially uncommon parameters or uncommon levels of parameters. Estimation of only revealed data, only stated data, and both revealed and stated data confirms this expectation. By combining the two types of data, estimation identifies several influential parameters, which generate insignificant or unexpected signs when using only the revealed data. In addition, regression analysis reveals that actual and hypothetical housing selections are guided by similar decision processes with respect to only certain parameters, such as the number of bedrooms, yet dissimilar processes with respect to other parameters, such as lot size. Future research should attempt to explore more deeply these divergent parameters. After adjusting for these divergent parameters, joint estimation of the combined data provides a test that finds that the hypothetical context of selecting a house from a prespecified choice set is similar to the market context of purchasing a home from the housing alternatives available in the real estate market.

Figure 1

Example of Conjoint Survey

Choice Set 1

Suppose you needed to leave your current home and were considering 3 houses to buy in Fairfield. The columns below describe these 3 housing options. The first house includes a water-based natural feature denoted by reference to the preceding photographs. The second house includes a land-based natural feature denoted by reference to the preceding photographs. (Each feature will remain natural for your entire time in the given house.) The third house includes neither feature.

Which house would you buy given your current financial situation?

House 1 __ House 2 __ House 3 __

	House 1	House 2	House 3
Natural Feature	Photo A	Photo G	Photo H
Number of Bedrooms	4	3	3
Number of Bathrooms	1	1	1
Internal Space (ft^2)	1500	1500	1500
Style	Colonial	Colonial	Ranch
Age (years)	new	70	new
Lot Size (acres)	0.2	0.6	0.6
Frequency of Flooding	never	never	never
Price	$ 250,000	$ 200,000	$ 600,000

Table 1

Attributes and Levels Included in Conjoint Analysis

Attribute	Levels	Attribute	Levels
Natural Feature	Long Island Sound	Age of House	0 years (new)
	Saltwater Marsh		40 years
	Freshwater Marsh		70 years
	River/Stream	Lot Size	0.2 acres
	Lake/Pond		0.6 acres
	Forest/Woods	Flooding	never
	Open Field/Park		every 100 years
	Backyard Lawns	Price	$ 200,000
Bedrooms	3		$ 250,000
	4		$ 350,000
Bathrooms	1		$ 600,000
	2	Style	Cape Cod
Interior Space	1,500 square feet		Colonial
	2,500 square feet		Ranch

Table 2

Characteristics of Sampled Households

Characteristic	Value	Percent (%)
Marital Status	Married	83.8
	Not Married	16.2
Dependent Children in House	Yes	56.2
	No	43.8
Size of Household	1	11.4
	2	29.5
	3	19.0
	4	28.6
	5	9.5
	6	1.9
Household Income	Less than $ 100,000	40.0
	$ 100,001 - $ 200,000	45.7
	More than $ 200,000	14.3

Table 3

Attributes of Sampled and Hypothetical Houses

Table 3.a. Average Values (standard deviations in parentheses)

Attribute	Revealed Data		Stated Data	
	Chosen Houses	Rejected Houses	Chosen Houses	Rejected Houses
Price ($)	303,924	291,937	282,123	351,925
	(166,320)	(165,102)	(112,867)	(160,383)
Age (years)	43.8	46.9	34.3	37.3
	(31.7)	(30.6)	(29.2)	(28.2)
Lot Size (acres)	0.681	0.886	0.353	0.330
	(0.855)	(3.022)	(0.195)	(0.187)
Interior space (ft^2)	1,964	1,955	1,910	1,798
	(1,019)	(916)	(492)	(457)
Bathrooms	2.30	2.34	1.36	1.31
	(1.030)	(0.998)	(0.481)	(0.461)
Bedrooms	3.32	3.40	3.34	3.33
	(0.935)	(0.909)	(0.472)	(0.469)

Table 3.b. Frequency Distributions (percent)

Attribute	Revealed Data		Stated Data	
	Chosen Houses	Rejected Houses	Chosen Houses	Rejected Houses
Natural Features				
Long Island Sound	2.9	1.9	12.3	4.8
Saltwater Marsh	5.7	1.3	8.4	8.1
Freshwater Marsh	0.9	1.5	8.5	6.8
River/Stream	8.6	2.9	9.7	5.8
Lake/Pond	0.0	1.3	4.7	2.7
Forest	21.0	17.8	24.2	21.5
Open Field / Park	2.8	2.8	14.6	11.1
Backyard Lawn	58.1	70.5	17.6	39.2
Style				
Cape Cod	15.2	28.2	30.2	32.4
Colonial	48.6	39.4	41.5	33.7
Ranch	21.0	14.9	28.3	33.9
Other	15.2	17.5	N/A	N/A
Flooding Risk				
Minimal	68.6	77.2	67.9	66.4
500-Year Flood	3.8	6.0	N/A	N/A
100-Year Flood	27.6	16.8	32.1	33.6
Census Tract				
Beach	30.5	23.2	N/A	N/A
Greenfield Hills	13.3	12.4	N/A	N/A
Other	56.2	64.4	N/A	N/A

Table 4

Pearson Correlation Coefficients for Revealed Data on Housing Purchases:

Selected Attributes

Attribute	Price	Age	Lot Size	Bedrooms	Bathrooms	Interior Space	Beach Census Tract
Price							
Age	- 0.21 ***						
Lot Size	0.21 ***	- 0.05					
Bedrooms	0.56 ***	0.01	0.18 ***				
Bathrooms	0.73 ***	0.17 ***	0.18 ***	0.66 ***			
Interior Space	0.78 ***	0.12 ***	0.20 ***	0.74 ***	0.80 ***		
Beach Census Tract	- 0.18 ***	- 0.11 **	- 0.13 ***	- 0.16 ***	- 0.21 ***	- 0.27 ***	
Greenfield Hills Census Tract	0.62 ***	0.17 ***	0.19 ***	0.31 ***	0.42 ***	0.47 ***	N/A

***, **, and * indicate statistical significance at levels of 1 %, 5 %, and 10 %, respectively, for test of nonzero values.

Table 5

Multinomial Logit Regression of Revealed Data

Variable [a]	Description	Coefficient Estimate	
Attributes			
Broad Natural Feature [b]	None (=0) versus	0	
	Water (=1)	2.977	***
		(1.175)	
	Land (=1)	0.524	
		(0.546)	
Water Feature	Freshwater Marsh (= -1) versus	- 4.966	
	Saltwater Marsh (=1)	2.184	***
		(0.821)	
	Long Island Sound (=1)	1.329	
		(0.866)	
	River/Stream (=1)	1.453	*
		(0.868)	
	Lake/Pond (=1) [c]	--	
Land Feature	Forest (=1) versus Field (= -1)	0.368	
		(0.529)	
Bedrooms	Number	- 0.148	
		(0.277)	
Bathrooms	Number	0.056	
		(0.315)	
Interior Space	1,000 ft^2	0.989	*
		(0.528)	

Style	Cape Cod (= -1) versus		
	Colonial (=1)	0.981	***
		(0.257)	
	Ranch (=1)	0.313	
		(0.260)	
	Other (=1)	0.932	***
		(0.261)	
Age	Years	- 0.008	
		(0.006)	
Lot Size	Acres	- 0.034	
		(0.142)	

Flooding	Minimal (= -1) versus	0.072	
	500-year Flood (=1)	- 0.672	
		(0.425)	
	100-year Flood (=1)	0.600	
		(0.425)	
Price	$ 1,000	- 0.019	**
		(0.010)	
Census Tract	Other (= -1) versus	- 0.591	
	Beach area (= 1)	- 0.319	
		(0.402)	
	Greenfield Hills (= 1)	0.910	**
		(0.401)	
Residual Quality [d]	$ 1	5.348	***
		(1.694)	
Household Characteristics Interacted with House Price			
Marital Status	Married (=1) versus Single (= -1)	0.003	
	[per $ 1,000]	(0.004)	
Children	Yes (=1) versus No (= -1)	- 0.0001	
	[per $ 1,000]	(0.002)	
Household Size	Number	0.003	
	[per $ 1,000]	(0.002)	
Income [e]	Low (= -1) versus	- 0.002	
	Medium (=1)	0.008	***
	[per $ 1,000]	(0.003)	
	High (=1)	0.010	***
	[per $ 1,000]	(0.004)	
Number of Observations	404		
Log-Likelihood	- 94.935		

Likelihood ratio statistic (χ^2)	95.41
McFadden's ρ^2	0.33

[a] Attributes with multiple levels are coded using effects codes, except as noted. Each level except the base level is represented by a column. Each column contains a "1" for the level represented by the column and a "-1" for the base level. The interpretation of these parameters is that the base level takes the utility level of the negative of the sum of the estimated coefficients and each other level takes the utility associated with the coefficient.

[b] Broad natural features are coded as 1,0 dummy variables.

[c] Observations involving lakes/ponds were deleted since no respondent chose these sites.

[d] Residuals from regression of the log values of house price on set of explanatory variables identical to discrete-choice hedonic analysis; residuals converted into dollar values.

[e] Low: < $ 100,000; Medium: $ 100,000 - $ 200,000; High: > $ 200,000.

Standard errors in parentheses. *,**,*** indicate statistical significance at levels of 10%, 5%, 1%, respectively.

Table 6

Multinomial Logit Regression of Stated Data

Variable [a]	Description	Coefficient Estimate	
Attributes			
Broad Natural Feature [b]	None (=0) versus	0	
	Water (=1)	1.838	***
		(0.348)	
	Land (=1)	1.189	***
		(0.244)	
Water Feature	Freshwater Marsh (= -1) versus	- 0.951	
	Saltwater Marsh (=1)	- 0.119	
		(0.115)	
	Long Island Sound (=1)	0.449	***
		(0.115)	
	River/Stream (=1)	0.244	**
		(0.120)	
	Lake/Pond (=1)	0.377	***
		(0.144)	
Land Feature	Forest (=1) versus Field (= -1)	0.172	**
		(0.084)	
Bedrooms	Number	0.077	
		(0.097)	
Bathrooms	Number	0.395	***
		(0.098)	

Interior Space	1,000 ft^2	0.668	***
		(0.100)	
Style	Cape Cod (= -1) versus	- 0.029	
	Colonial (=1)	0.149	***
		(0.058)	
	Ranch (=1)	- 0.120	**
		(0.062)	
	Other (=1)	N/A	
Age	Years	- 0.004	***
		(0.002)	
Lot Size	Acres	0.869	***
		(0.244)	

Flooding	Minimal (= -1) versus	0.065	
	100-year Flood (=1)	- 0.065	
		(0.051)	
Price	$ 1,000	- 0.0017	
		(0.0018)	
Household Characteristics Interacted with House Price			
Marital Status	Married (=1) versus Single (= -1)	- 0.001	*
	[per $ 1,000]	(0.0007)	
Children	Yes (=1) versus No (= -1)	- 0.0002	
	[per $ 1,000]	(0.0006)	
Household Size	Number	0.0004	
	[per $ 1,000]	(0.0006)	
Income [c]	Low (= -1) versus	- 0.005	
	Medium (=1)	0.001	***
	[per $ 1,000]	(0.0005)	
	High (=1)	0.004	***
	[per $ 1,000]	(0.0006)	
Number of Observations	2,727		
Log-Likelihood	- 791.043		
Likelihood Ratio Statistic (χ^2)	407.253		
McFadden's ρ^2	0.20		

[a] Attributes with multiple levels are coded using effects codes, except as noted. Each level except the base level is represented by a column. Each column contains a "1" for the level represented by the column and a "-1" for the base level. The interpretation of these parameters is that the base level takes the utility level of the negative of the sum of the estimated coefficients and each other level takes the utility associated with the coefficient.

[b] Broad natural features are coded as 1,0 dummy alternative-specific constants.

[c] Low: < $ 100,000; Medium: $ 100,000 - $ 200,000; High: > $ 200,000.

Standard errors in parentheses. *,**,*** indicate statistical significance at levels of 10%, 5%, 1%, respectively.

Table 7

Classification of Parameters for Compatibility of Revealed and Stated Data

Compatible Parameters:

Water-Based Natural Features

Land-Based Natural Features

Long Island Sound natural feature

River/Stream natural feature

Forest natural feature

Bedrooms

Interior Space

Age

Presence of Dependent Children- Interaction with Price

Medium Income - Interaction with Price

High Income - Interaction with Price

Incompatible Parameters:

Price

Marital Status - Interaction with Price

Household Size - Interaction with Price

Saltwater Marsh natural feature

Colonial Style

Ranch Style

Lot Size

100-year Flooding Risk

Parameters Not Common to Both Data Sets:

Lake/Pond natural feature

“Other” Style

Beach Census Tract

Greenfields Census Tract

500-year Flooding Risk

Hedonic Price Residual

Table 8

Multinomial Logit Regression of Combined Revealed and Stated Data

Variable [a,b]	Description	Coefficient Estimate [c]	
		Revealed Data	Stated Data
Attributes			
Broad Natural			
Feature [d]	None (=0) versus		0
	Water (=1)	1.714 ***	
		(0.311)	
	Land (=1)	1.014 ***	
		(0.198)	
Water Feature	Freshwater Marsh (= -1) versus	- 2.429	- 0.969
	Saltwater Marsh (=1)	1.333 ***	- 0.127
		(0.521)	(0.117)
	Long Island Sound (=1)	0.455 ***	
		(0.117)	
	River/Stream (=1)	0.260 **	
		(0.123)	
	Lake/Pond (=1)	0.381 ***	
		(0.145)	
Land Feature	Forest (=1) versus Field (= -1)	0.179 **	
		(0.086)	
Bedrooms	Number	0.022	
		(0.092)	

Bathrooms	Number			0.332	***		
				(0.095)			
Interior Space	1,000 ft^2			0.681	***		
				(0.101)			
Style	Cape Cod (= -1) versus	- 1.668				- 0.094	
	Colonial (=1)	0.718	***			0.156	***
		(0.217)				(0.060)	
	Ranch (=1)	0.737	***			- 0.119	*
		(0.238)				(0.064)	
	Other (=1)			0.213			
				(0.243)			
Age	Years			- 0.004	***		
				(0.002)			
Lot Size	Acres	- 0.084				0.890	***
		(0.135)				(0.254)	
Flooding	Minimal (= -1) versus	- 0.211				0.525	
	500-Year Flood (=1)			- 0.454			
				(0.332)			
	100-year Flood (=1)	0.665	*			- 0.071	
		(0.400)				(0.535)	
Price	$ 1,000	- 0.026	***			- 0.001	
		(0.005)				(0.002)	
	Household Characteristics Interacted with House Price						
Marital Status	Married (=1) versus Single (= -1)	0.007	**			- 0.002	**
	[per $ 1,000]	(0.003)				(0.001)	
Children	Yes (=1) versus No (= -1)			- 0.0003			
	[per $ 1,000]			(0.0006)			

Household Size	Number	0.0033	***			0.0002	
	[per $ 1,000]	(0.001)				(0.001)	
Income [e]	Low (= -1) versus			- 0.0066			
	Medium (=1)			0.0019	***		
	[per $ 1,000]			(0.001)			
	High (=1)			0.0047	***		
	[per $ 1,000]			(0.001)			
Number of Observations	3,131						
Log-Likelihood	- 931.702						
Likelihood Ratio Statistic (χ^2)	481.549						
McFadden's ρ^2	0.21						

[a] Attributes with multiple levels are coded using effects codes, except as noted. Each level except the base level is represented by a column. Each column contains a "1" for the level represented by the column and a "-1" for the remaining levels. The interpretation of these parameters is that the base level takes the utility level of the negative of the sum of the estimated coefficients and each other level takes the utility associated with the coefficient.

[b] Tables shows only variables common to both the stated and revealed data. The regression additionally includes the uncommon variables.

[c] Parameters with only one reported value are constrained to be equal across the two data sets.

[d] Broad natural features are coded as 1,0 dummy variables.

[e] Low: < $ 100,000; Medium: $ 100,000 - $ 200,000; High: > $ 200,000.

Standard errors in parentheses.

*,**,*** indicate statistical significance at levels of 10%, 5%, 1%, respectively.

Stated data is re-scaled by a factor of 0.85.

REFERENCES

Adamowicz, Wiktor, J. Louviere, and M. Williams (1994), "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, v. 26, n. 3, pg. 271-292.

Adamowicz, Wiktor, Joffre Swait, P. Boxall, Jordan Louviere, and M. Williams (1997), "Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation," *Journal of Environmental Economics and Management*, v. 32, pg. 65-84.

Anas, A. (1982), *Residential Location Markets and Urban Transportation*, Academic Press: NYC.

Anas, A. and C. Chu (1984), "Discrete Choice Models and the Housing Price and Travel to Work Elasticities of Location Demand," *Journal of Urban Economics*, v. 15, pg. 107-123.

Bartik, Timothy (1987), "The Estimation of Demand Parameters in Hedonic Price Models," *Journal of Political Economy*, v. 95, n. 1, pg. 81-88.

Ben-Akiva, Mosha and Steve Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press: Cambridge, MA.

Boehm, T.P., H.W. Herzog, and A.M. Schlottmann (1991), "Intra-urban Mobility, Migration, and Tenure Choice," *Review of Economics and Statistics*, v. 73, pg. 59-68.

Brown, Gardner M. Jr., and Henry Pollakowski (1977), "Economic Valuation of Shoreline," *Review of Economics and Statistics*, v. 59, n. 3, pg. 272-278.

Bunch, D.S., J.J. Louviere, and D.A. Anderson (1992), "A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes," unpublished working paper, Graduate School of Management, University of California, Davis.

Cameron, Trudy (1992), "Combining Contingent Valuation and Travel Cost for the Valuation of Nonmarket Goods," *Land Economics*, v. 68, pg. 302-317.

Chapman, David, W. Michael Hanemann, and Barbara Kanninen (1996), "Non-Market Valuation Using Contingent Behavior: Model Specification and Consistency Tests," presented at 1996 Association of Environmental and Resource Economists Summer Workshop, Lake Tahoe, CA, June 2-4, 1996.

Cropper, Maureen L., Leland B. Deck, and Kenneth E. McConnell (1988), "On the Choice of Functional Form for Hedonic Price Functions," *Review of Economics and Statistics*, v. 70, n. 4, pg. 668-675.

Cropper, Maureen L., Leland B. Deck, Nalin Kishor, and Kenneth E. McConnell (1993), "Valuing Product Attributes Using Single Market Data: A Comparison of Hedonic and Discrete Choice Approaches," *Review of Economics and Statistics*, v. 75, pg. 225-232.

Ellickson, Bryan (1977), "Economic Analysis of Urban Housing Markets: A New Approach," R-2024 NSF (Rand Corporation, Santa Monica, CA).

Ellickson, Bryan (1981), "An Alternative Test of the Hedonic Theory of Housing Markets," *Journal of Urban Economics*, v. 9, pg. 56-79.

Frech, H.E. III, and Ronald Lafferty (1984), "The Effect of the California Coastal Commission on Housing Prices," *Journal of Urban Economics*, v. 16, pg. 105-123.

Freeman, A. Myrick (1991), "Factorial Survey Methods and Willingness to Pay for Housing Characteristics: Comment," *Journal of Environmental Economics and Management*, v. 20, pg. 92-96.

Freeman, A. Myrick (1993), *The Measurement of Environmental and Resource Values: Theory and Methods*, Resources for the Future: Washington, DC.

Friedman, J. (1981), "A Conditional Logit Model of the Role of Local Public Services in Residential Choice," *Urban Studies*, v. 18, pg. 347-358.

Goodman, Allen (1989), "Identifying Willingness-to-Pay for Heterogeneous Goods with Factorial Survey Methods," *Journal of Environmental Economics and Management*, v. 16, pg. 58-79.

Graves, Phil, James Murdoch, Mark Thayer, and Don Waldman (1988), "The Robustness of Hedonic Price Estimation: Urban Air Quality," *Land Economics*, v. 64, pg. 220-33.

Greene, William (1993), *Econometric Analysis*, Prentice-Hall: Englewood Cliffs, NJ.

Gross, David, C.F. Sirmans, and John Benjamin (1990), "An Empirical Evaluation of the Probabilistic Bid-Rent Model: The Case of Homogeneous Households," *Regional Science and Urban Economics*, v. 20, pg. 103-110.

Huang, Ju-Chin, and Timothy Haab, and John Whitehead (1997), "Willingness to Pay for Quality Improvements: Should Revealed and Stated Preference Data be Combined?" *Journal of Environmental Economics and Management*, v. 34, pg. 240-255.

King, A.T. (1980), "An Economic Model of Tenure Choice and Demand for Housing in a Joint Decision," *Journal of Public Economics*, v. 14, pg. 137-159.

Lerman, Steve (1977), "Location, Housing, Automobile Ownership, and the Mode to Work: A Joint Choice Model," *Transportation Research Board Record*, v. 610, pg. 6-11.

Longley, P.A. (1984), "Comparing Discrete Choice Models: Some Housing Market Examples," in *Discrete Choice Models in Regional Science*, ed., D.E. Pitfield, Pion Limited: London, pg. 163-180.

Louviere, Jordan and D.A. Hensher (1982), "On the Design and Analysis of Simulated Travel Choice or Allocation Experiments in Travel Choice Modeling," *Transportation Research Record*, v. 890, pg. 11-17.

Louviere, Jordan and D.A. Hensher (1983), "Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event," *Journal of Consumer Research*, v. 10, pg. 348-361.

Louviere, Jordan and G.G. Woodworth (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, v. 20, pg. 350-367.

Louviere, Jordan (1988), "Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments, and External Validity," *Journal of Transport Economics and Policy*, v. 22, n. 1, pg. 93-119.

Louviere, Jordan (1996), "Combining Revealed and Stated Preference Data: The Rescaling Revolution," presented at 1996 Association of Environmental and Resource Economists

Summer Workshop, Lake Tahoe, CA, June 2-4, 1996.

McFadden, D. (1975), “On Independence, Structure and Simultaneity in Transportation Demand Analysis,” Working Paper 7511, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley.

McFadden, D. (1978), “Modeling the Choice of Residential Location,” in *Spatial Interaction Theory and Planning Models*, eds., Anders Karlqvist et. al., North-Holland: Amsterdam, pg. 75-96.

Mitchell, Robert Cameron and Richard T. Carson (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Methods*, Resources for the Future: Wash, DC.

Nechbya, Thomas and Robert Strauss (1988), “Community Choice and Local Public Services: A Discrete Choice Approach,” *Regional Science and Urban Economics*, v. 28, pg. 51-73.

Palmquist, Raymond (1984), “Estimating the Demand for Characteristics of Housing,” *Review of Economics and Statistics*, v. 64, n. 3, pg. 394-404.

Palmquist, Raymond (1992), "Valuing Localized Externalities," *Journal of Urban Economics*, v. 31, n. 1, pg. 59-68.

Parsons, George and Mary Jo Kealy (1992), “Randomly Drawn Opportunity Sets in a Random Utility

Model of Lake Recreation," *Land Economics*, v. 68, pg. 93-106.

Quigley, John (1976), "Housing Demand in the Short Run: An Analysis of Polytomous Choice," *Explorations in Economic Research*, v. 3, pg. 76-102.

Quigley, John (1985), "Consumer Choice of Dwelling, Neighborhood, and Public Services," *Regional Science and Urban Economics*, v. 15, pg. 41-63.

Rosen, Sherwin (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Perfect Competition," *Journal of Political Economy*, v. 82, n. 1, pg. 34-55.

Steadman, Geoffrey (1996), "Multiple Use Management Plan for Coastal Open Space: Volume II," prepared for Fairfield Conservation Commission, April.

Swait, Joffre and Wiktor Adamowicz (1996), "The Effect of Choice Complexity on Random Utility Models: An Application to Combined Stated and Revealed Preference Models or Tough Choices: Contribution or Confusion?" presented at 1996 Association of Environmental and Resource Economists Summer Workshop, Lake Tahoe, CA, June 2-4, 1996.

Swait, Joffre and J. Louviere (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models," *Journal of Marketing Research*, v. 30, pg. 305-314.

Swait, Joffre, Jordan Louviere, and M. Williams (1994), "A Sequential Approach to Exploiting the Combined Strengths of SP and RP Data: Application to Freight Shipper Choices," *Transportation*, v. 21, pg. 135-152.

Timmermans, H., A. Borgers, and J. van Dijk, and H. Oppewal (1992), "Residential Choice Behavior of Dual-Earner Households: A Decompositional Joint Choice Model," *Environment and Planning A*, v. 24, pg. 517-533.

Timmermans, H. and L. van Noortwijk (1995), "Context Dependencies in Housing Choice Behavior," *Environment and Planning A*, v. 27, pg. 181-192.

Train, Kenneth, and Terry Atherton (1995), "Rebates, Loans, and Customers' Choices of Appliance Efficiency Level: Combining Stated and Revealed Preference Data," *Energy Journal*, v. 16, n. 1, pg. 55-69.

William, Roberton Jr. (1979), "A Logit Model of Demand for Neighborhood," in *The Economics of Neighborhood*, ed., David Segal, Academic Press: New York, pg. 17-42.