

Meta Analysis in Model Implementation: Choice Sets and the Valuation of Air Quality Improvements

H. Spencer Banzhaf and V. Kerry Smith

November 2003 • Discussion Paper 03-61



RESOURCES
FOR THE FUTURE

Resources for the Future
1616 P Street, NW
Washington, D.C. 20036
Telephone: 202-328-5000
Fax: 202-939-3460
Internet: <http://www.rff.org>

© 2003 Resources for the Future. All rights reserved. No portion of this paper may be reproduced without permission of the authors.

Discussion papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review or editorial treatment.

Meta Analysis in Model Implementation: Choice Sets and the Valuation of Air Quality Improvements

H. Spencer Banzhaf and V. Kerry Smith

Abstract

We document the sensitivity of welfare estimates derived from discrete choice models to assumptions about the choice set. Such assumptions can affect welfare estimates through both the estimated parameters of the model and, conditional on the parameters, the substitution among alternatives. Our analysis involves estimates of the benefits of air quality improvements in Los Angeles based on discrete choices of neighborhood and housing. We further illustrate the use of meta analysis to document and summarize voluminous information derived from repeated sensitivity analyses.

Key Words: Meta analysis, random utility model, choice set, air quality, housing.

JEL Classification Numbers: C15, Q25, R21

Contents

1. Introduction.....	1
2. Modeling Housing Choice and the Choice Set	5
3. Empirical Illustration	7
4. Digression on the Efficient Sample Size.....	11
5. Results	12
6. The Effects of Choice Sets in Other Applications	15
7. Summary and Conclusions.....	17
References.....	19

Meta Analysis in Model Implementation: Choice Sets and the Valuation of Air Quality Improvements

H. Spencer Banzhaf and V. Kerry Smith*

1. Introduction

Applying economic theory to actual problems invariably requires decisions about how best to translate a conceptual framework into a specific econometric model. Judgments about the appropriate functional form to be used are the most obvious example, but by no means the only. Such judgments can greatly influence the conclusions drawn from analysis, yet theory may not be any guide. Although understanding the sensitivity of results is important for assessing empirical work and in many cases may be a separate source of insight, the role of such judgments rarely are documented in published manuscripts.¹ Full disclosure of the background research for an empirical article requires the reporting of sensitivity analyses, evaluations of the robustness of conclusions, and so forth, yet space limitations, together with a desire to avoid the appearance of “data mining,” have constrained the presentation of this type of detail. Consequently, at present, most of the insights gained from model development remain with authors. Fellow practitioners and “consumers” of applied research in the business and policy worlds lose out on these insights and implications.²

* H. Spencer Banzhaf is a Fellow at Resources for the Future and V. Kerry Smith is a University Distinguished Professor in the Department of Agricultural and Resource Economics, North Carolina State University and a University Fellow at Resources for the Future. Partial support for Smith’s research was provided by the U.S. Environmental Protection Agency under grant numbers R 828103 and R 82950801. Earlier versions of this paper were presented at MASTERPOINT, *Meta Analysis in Economics, An International Colloquium* at the Free University in Amsterdam, and Camp Resources. Thanks are due to Raymond Florax and other conference participants for helpful comments.

¹ Ellison’s (2002) recent detailed empirical analysis of the slowdown in the economic publishing process finds no support for increasing attention over time to quality as measured by robustness checks, discussions of related literature, etc. (p. 988). This finding implies that, to date, decisions to publish have not given greater weight to evaluating the judgments made in empirical studies.

² For example, in the policy world, the Office of Management and Budget has recently proposed guidelines for conducting benefit-cost analyses in rulemaking. These guidelines require extensive documentation of measures used in benefit or cost transfers (see *Appendix C, OMB Draft Guidelines for the Conduct of Regulatory Analyses and the Format of Accounting Statements*, Federal Register, 68 (22), Monday, February 3, 2003, pp. 5513-5527.

Meta analysis can concisely document the effects of such judgments. Typically, meta analyses synthesize results from different empirical studies by different researchers. In contrast, we suggest that individual researchers (or teams) may also use meta analysis to summarize the influence of alternative modeling judgments in their own work. When a large number of models generate too much information to document as separate results, or to process cognitively, meta analysis can both simplify the presentation and highlight the statistical signals.

We illustrate our proposal with an application to random utility models of household choices for heterogeneous housing.³ Our particular focus is on the use of these models to estimate the economic value of reducing ozone concentrations (e.g., Chattopadhyay 2000, Banzhaf 2002). However, they have also been used for many other applications, including models of demand for differentiated produces (Berry, Levinsohn, and Pakes 1995, 2003; Nevo 2003) and more recently models of endogenous neighborhood formation (Brock and Durlauf 2001, 2002; Bayer 2000).

Estimating these models requires sufficient intra-neighborhood variation in house-level characteristics (e.g., square footage) and inter-neighborhood variation in regional characteristics (e.g., air and school quality) to identify the relevant parameters. The relevant variation is over the set of alternative neighborhoods and houses among which each household can choose (the choice set). Applying this model first requires a concrete judgment about this choice set: which alternatives in the data were actually available to the household? Over which was the maximizing choice made? For example, our application involves over 300,000 housing sales in the Los Angeles area over a four-year period. Clearly, every household did not choose among all the houses that were for sale in the entire five-county region and over the entire time period.⁴

³ See Palmquist (2003) for a review placing this work in the context of other models such as hedonic pricing.

⁴ A related question is what set of alternatives households actively “considered.” Known as consideration sets, these sets may be determined by supplemental information from surveys. See, for example, Peters, Adamowicz, and Boxall (1995) and Ben-Akiva and Boccara (1995). In a similar vein, instead of using survey data, Haab and Hicks (1997) assume in an application to outdoor recreation that, for each choice occasion, the choice set consists of those sites that households were observed to have chosen on at least one other occasion (since this is evidence that the site must be known to the household). A potential criticism of this general approach is that alternatives outside the consideration set may well be known to the household, but may be less preferred. Consequently, the modeled set may not reflect all alternatives that are truly available, but only those considered after others have been eliminated in a first-stage cognitive process. Ben-Akiva and Boccara (1995) and Horowitz and Louviere (1995) suggest models that exploit consideration sets for additional information about underlying preferences. In any case, the required supplemental information is not available in our application.

Moreover, because they generally assume that households maximize over large choice sets, empirical applications often rely on random sampling rules during estimation. These rules involve sampling a subset of alternatives from the overall choice set, that is, the household is modeled as if its observed choice was selected from a smaller set. McFadden (1978) demonstrated that such sampling was consistent so long as the sampling rule satisfies the so-called positive conditioning property, or, what is usually used in practice, a stronger property known as uniform conditioning.⁵ The Independence of Irrelevant Alternatives (IIA) property of the simple conditional logit model (conditional IIA property of nested logit models) assures that preference parameters can be estimated consistently from the sampled choice sets.⁶

Judgments made at both stages—the specification of the “true” overall choice set and the scheme for sampling from this choice set during estimation—can affect the welfare estimates.⁷ First, they may influence the estimated parameters of the structural model. Equally important, even conditioned on any given set of parameters, the specification of the true choice set will affect welfare measures by restricting the set of relevant substitutes. In other words, the value of a price or quality change hypothesized for one or more alternatives will be a function of both the alternatives affected and the others available.⁸

Our application involves housing in a five-county area around Los Angeles from 1989 to 1994. Realistically, each of the many houses observed to be on the market could not possibly have been in the choice set of each household. We evaluate judgments about the true choice set along three dimensions. The first is time: over what period of time do households actively

⁵ Positive conditioning requires that the probability of drawing the sample is always non-zero, regardless of which alternative in the sample is actually chosen. The uniform conditioning property, a sufficient condition, is more tractable. It requires that the probability of drawing the sample is the same regardless of which alternative is actually chosen.

⁶ IIA restricts the pattern of substitution across alternatives so that the ratio of choice probabilities between any two alternatives is independent of the other alternatives in the choice set. In other words, the elasticity of the choice probabilities for any two alternatives with respect to a third alternative are the same.

⁷ Parsons and Hauber (1998) and Parsons and Kealy (1992) demonstrated sensitivity of results to each respective type of judgment in an application to choices of sites for outdoor recreation and household values for water quality.

⁸ In random utility models, the error is interpreted as a part of unobserved preference heterogeneity, implying it should be integrated into welfare measures. Thus, willingness to pay is defined implicitly by equating the maximum of the conditional indirect utility functions before and after the policy change. These functions are random variables and, for the central case of errors that are distributed according to the Type I extreme value distribution, will be functions of all elements in the choice set. To our knowledge, this point was first recognized by Hanemann (1978) and used as an issue in defining recreation choice sets in Kaoru and Smith (1990).

search? The second is geographic space: over what area do households search? And the third is available resources for housing purchases: over what price range do households search?

We compare 24 different specifications of the true choice set, using “boundaries” in each dimension. For each boundary of the true choice set, we draw samples for estimation from the same set, and from proper subsets defined by tighter boundaries in a way that satisfies McFadden's uniform conditioning property. We draw five samples from each possible combination of boundaries, each with 40,000 observations and 15 choice alternatives. On each sample, we then estimate conditional logit models using two functional specifications, the natural log and the square root functions of the observed housing attributes, and compute welfare measures using the larger “true” choice set.⁹ We then summarize the importance of these alternative judgments with a meta analysis. The welfare estimates for each case are the dependent variables in the meta analysis, with indicator variables for the true choice set boundaries and for the sampling boundaries as the explanatory variables.

Our meta analysis indicates that the time window dimension of the choice set has a significant, but small, effect on estimates of the willingness to pay for an improvement in ozone (the most salient pollutant in the Los Angeles area). Geographic boundaries for the true choice set also appear to influence the results, but not the sampling from the choice set. Finally, the budgetary dimension of the choice set is quite important to both estimation and welfare measurement with these models.

Section 2 provides some background describing why these implementation issues might be expected to influence results. It also considers how the choice set influences estimation and welfare measurement in a random utility framework. In Sections 3 and 4, we outline our empirical strategy in more detail, and in Section 5 present our meta analysis. Section 6 draws from these results to gain insight into, and to comment on, past efforts to assess the sensitivity of random utility models to choice sets. We close with some general discussions about using meta analysis in other situations to gauge the robustness of models to various implementation decisions.

⁹ These specifications have been widely used in the literature. Using simulated data, Cropper et al. 1993 found that the square root had the smallest error in estimating marginal values of attributes, regardless of the true form. The true specification was better only under the correct form. Otherwise, the average error was smaller with the square root than for the true specification when some attributes were not observed or replaced with proxies.

2. Modeling Housing Choice and the Choice Set

The conceptual model assumes that households select a house that yields the greatest utility, given their income and given the price and the attributes of the home. A stochastic indirect utility function describes the choice model for estimation as in equation (1):

$$V_{ij} = \phi(z_j, m_i - p_j^h) + \varepsilon_{ij}, \quad (1)$$

where V_{ij} is the utility of house j for household i , z_j is the vector of attributes at house j , m_i is the income of household i , p_j^h is the annualized (or rental) price of house j , and ε_{ij} is an idiosyncratic taste shock realized by household i for house j . The price of all other goods is normalized to one.

The probability household i will select house k over j is then given by the probability that $V_{ik} > V_{ij}$, or

$$\begin{aligned} \text{prob}(z = z_k | i) = \\ \text{prob} \left\{ \phi(z_k, m_i - p_k^h) - \phi(z_j, m_i - p_j^h) > \varepsilon_{ij} - \varepsilon_{ik} \right\} \quad \forall j \neq k. \end{aligned} \quad (2)$$

If all the ε 's are assumed to be independent and identically distributed by the Type I extreme value distribution, then the analytical form for probability defined in equation (2) is given by equation (3):

$$\text{prob}(z = z_k | i) = \frac{\exp[\phi(z_k, m_i - h(z_k))]}{\sum_{j \in J_i} \exp[\phi(z_j, m_i - h(z_j))]} \quad (3)$$

Note that the choice set for household i , J_i , enters through the denominator.

Judgments about the choice set come into play at two distinct points. The first point is estimation of the parameters in (1) using the likelihood function derived from (3). As noted previously, McFadden has demonstrated that, when IIA holds, to estimate the parameters, J may be replaced with any set J' ($J' \subseteq J$) that satisfies the uniform conditioning property. Random sampling reduces the complexity caused by the scale of large problems. However, there is an

important economic question that should not be overlooked. To define a sampling process satisfying uniform conditioning, the analyst must still specify the “true” choice set J from which the sample is to be drawn.¹⁰

Including alternatives that are not actually in the choice set can bias the estimated parameters. Accordingly, a conservative sampling strategy would draw the alternatives most likely to be in the true choice set. In our application, L is the set of all houses on the market. For example, suppose that the universe of alternatives that may be in the choice set is the set L , $J \subseteq L$. If L can be partitioned into two sets, $L1$ and $L2$, and if the elements of $L1$ are believed more likely to be in J than the elements of $L2$, then it would be preferable to estimate the model with $L1$ (or a subset of $L1$) than the entire set L , or certainly than $L2$. For example, in our application to housing choice, the length of time a household searches for a house is unobserved. Most transaction-based databases record only that a house was purchased on a given date, say, June 1st. Surely, a house that was on the market on May 31st was in the household's choice set (assuming it satisfies other criteria). But what about a house that was on the market January 1st? In this situation, the targeted approach to specifying the population for sampling would regard inclusion of the first housing alternative as more plausible than the second.

Unfortunately, this analysis may seem deceptively simple. Consider the boundary in a dimension other than time, the price of housing. Less expensive houses might be more likely to be in a household's choice set than more expensive ones. While, in theory, sampling based on such observables is unbiased under the correct specification, in practice it can cause difficulties. As discussed below, we have difficulty identifying the price parameter under the more conservative boundaries in this dimension.

The second point where judgments about the choice set come into play is in welfare computations. The conventional Hicksian willingness to pay (WTP) measure is based on the maximum of the stochastic indirect utility functions with and without the change being evaluated, as in equation (4):

$$\text{Max}_{\{j \in J_i\}} V_{ij} \left(z_j, m_i - p_j^h, \varepsilon_{ij} \right) = \text{Max}_{\{j \in J_i\}} V_{ij} \left(z_j^*, m_i - p_j^h - WTP, \varepsilon_{ij} \right) \quad (4)$$

¹⁰ This issue does not arise as directly in alternatives to the random utility model, such as the hedonic pricing model. Conventional interpretations of the hedonic price function suggest it is a statistical approximation of the market equilibrium implied by a matching of buyers and sellers. Concerns about choice set in this context have a parallel in defining the extent of the market and its implications for the equality of marginal (implicit) prices of a good's attributes over time and space.

where z^* is the value of the attributes after some policy shock (for example, an improvement in air quality). With the Type I extreme value distribution, the distribution of these random variables (i.e., the maximums of V_{ij} with and without the change) are also Type I extreme values, and their location parameters are functions of the location parameters for *all alternatives* in the true choice set.¹¹ Sampling does *not* resolve this issue. Consequently, judgments must be made about the composition of the true choice set J .

3. Empirical Illustration

We illustrate how sensitivity analysis of such modeling judgments may be summarized using meta analysis techniques. Our application considers alternative definitions of the choice set in random utility models of housing choice in Los Angeles from 1989 to 1994. The model's objective is to estimate the benefits of air quality improvements in this area.

The actual prices and structural characteristics are observed for houses in Orange, Los Angeles, Riverside, San Bernardino, and Ventura counties. These data were obtained from Transamerica Intellitech, a market research firm. About half of the transactions recorded housing prices. After deleting about 10 percent of the remaining observations as outliers or because of inconsistent values, 319,641 observations remained for analysis. Each observation is associated with a unique house. These data form the universe of alternatives that potentially are in each household's choice set, L .

The data include the size of the lot, the area of the house in square feet, the number of bathrooms and bedrooms, the presence of a fireplace, the presence of a swimming pool, and the age of the house. In addition, the location of each house is specified by its latitude and longitude. This spatial information allows each record to be linked to separate data on location-specific amenities, including schooling, crime, proximity to the coast, and air quality, the public good of interest in our application.

Our measure of air quality is the ambient concentration of ozone, defined as the number of days without an exceedence of the national ozone standard. This measure has the advantage of coinciding with the information communicated to residents as smog alerts (e.g., on the *LA Times* weather page) and of being consistent with ozone's acute health effects. There is a good

¹¹ See Ben-Akiva and Lerman (1985) for an overview of their properties.

deal of variability and precision in these measures, as Los Angeles is one of the most densely monitored regions in the world, with an average of 50 monitors available each year in the study area, plus monitors in neighboring counties.¹² Ozone was imputed for each house using the nearest monitor, with the median distance to a monitor being 4.5 miles.

Educational amenities include the average scores on standard math tests conducted in the tenth grade and the teacher-student ratios in each school district, and were obtained from the National Center for Education Statistics. Crime rates were obtained from the California Department of Justice. Proximity to the coast is captured with an indicator variable for being within one mile. Finally, household income data were imputed from the U.S. Census averages at the block-group level (with a median of 514 households and eight annual transactions per block-group). Table 1 summarizes the means of all these variables by county.

Choice sets were defined by specifying boundaries in three dimensions. The first dimension is geographic space, as households may only consider houses in a certain area. These implied restrictions can arise due to a desire to reside in a particular neighborhood or because of commuting distance. In our application to Los Angeles, we define the spatial boundaries as either the county of actual residence or the entire LA area. Smaller areas were not used, as they would not provide the variation required to estimate the value of the spatially delineated local public goods.

The second dimension is the budget share allocated to housing. Choice sets are defined using budget shares for annualized housing prices less than or equal to 100 percent, 52 percent, and 44 percent of imputed annual income. The last two boundaries represent the estimated 95th and 90th percentile, respectively, of the empirical distribution in the Los Angeles housing data after matching houses to income data from the U.S. census.¹³

¹² Information on these data can be obtained from the California Air Resources Board's web page at www.arb.ca.gov/aqd/aqd.htm. Information is also available from the U.S. EPA at www.epa.gov/airsweb.

¹³ According to the Bureau of Labor Statistics, the mean budget share of housing in the United States in 1990 was 28 percent. The estimated mean in these data is 30 percent. This dimension of the boundary is especially complex for another reason: Most applications provide little information about the wealth of households and how it influences the financing of new home purchases. Households that previously owned homes under most recent income tax regimes were induced by the statutes to “roll over” capital gains into their new primary residence. Thus, the relationship between affordable rents (with current income) and annualized price depends on a household’s portfolio of assets and how much each is willing to hold in the house it occupies.

The third dimension is the time households and houses are assumed to be “in the market.” According to the California Association of Realtors (2000), the mean time a typical house was on the market during our sample period was about two and one-half months. Assuming the median for our sample was close to this value, a time-window of plus or minus two and one-half months around the date of the household’s actual choice, or a window of five months, would provide a reasonable temporal choice set.¹⁴ Alternative temporal windows of one, three, and seven months were also used to define the choice set. Increasing the time window increases the size of the choice set without any clear correlation with other economic variables.

The four temporal windows, three budget constraints, and two geographic regions provide a total of 24 “true” choice sets. For each type of choice set, we draw five independent samples and estimate the preference functions with each choice set. For each of these models, we then simulate the exact compensating variation for a nonmarginal improvement in ozone for the entire choice set defined by the boundaries used in estimation, *and for each of the choice set types of which it is a subset*. The logic for this strategy stems from the fact that the positive conditioning property allows a model, with any given choice set boundary, to be estimated with any smaller choice set meeting the uniform conditioning property. The reverse of this logic is that any model estimated with a sample from a given boundary is consistent with the true boundary being larger. For example, if the model is estimated with a choice set sampled from houses in the same county as the chosen house that represent a budget share under 52 percent, and that are within a five-month time window of the house, then we calculate welfare for the following complete choice sets:

- 5 months, 52 percent of income, entire LA market;
- 5 months, 100 percent of income, same county;
- 5 months, 100 percent of income, entire LA market;
- 7 months, 52 percent of income, same county;
- 7 months, 52 percent of income, entire LA market;
- 7 months, 100 percent of income, same county; and
- 7 months, 100 percent of income, entire LA market.

¹⁴ Ideally, one would use the average time on the market of houses to extend the window forward, as here, but use the average search time of households to extend the window backward.

Each household's conditional utility function is specified in two alternative forms, both nonlinear in income. Specifically, $\phi_{ij}(\cdot)$ represents either the natural log or the square root function. Thus, we estimate the following equation with $f(\cdot)$ representing logs or square roots:

$$V_{ij} = \alpha f(m_i - p_j^h) + \beta f(q_j) + \sum_r \gamma_r f(z_{rj}) + \sum_s \delta_s z_{sj} + \varepsilon_{ij} \quad (5)$$

Here, q is air quality and z is now the vector of remaining attributes listed in Table 1, r indexing the continuous attributes and s the discrete attributes represented by dummy variables. Housing asset prices p^h are annualized using a factor of 0.116 from Poterba (1992).

Our policy intervention improves air quality by decreasing the number of exceedences by five to 25 days. The improvement at each house is drawn from a uniform distribution, giving a mean 15-day improvement (the average improvement in Los Angeles from 1989 to 1994). The policy introduces heterogeneity in the improvement in order to create the greatest possible advantage to households with choice sets involving larger time windows. With heterogeneity, larger choice sets create better opportunities to select a house with a large improvement in air quality. For each case, the exact compensating variation is calculated for 200 households purchasing in time windows centered around June 1989. Each household is characterized by draws from the empirical income distribution in Los Angeles and from a random variable determining the household's "home county."¹⁵

With each of the 24 choice sets J , welfare measures are calculated as the expected willingness to pay: the payment required after the air quality improvement to maintain utility at its baseline level. WTP is defined implicitly as in Equation (4). It equates realized utility in the comparison scenario with that in the reference scenario, *at the house that is chosen after the payment is made*. This measure assumes that households may freely re-optimize their choice of housing location after the change in public goods. The expectation is taken over the distribution of the error term. With nonlinear income effects, this measure does not have a closed-form solution.¹⁶ As a result, draws are made from the Type I extreme value distribution for each household characterized by income and home county and the WTP is estimated to within one

¹⁵ A log-normal income distribution estimated from the quantiles reported by the U.S. census was used for the former, while a discrete distribution matching 1990 population was used for the latter.

¹⁶ See McFadden (1999) and Herriges and Kling (1999).

dollar using a numerical bisection routine. This process is repeated over a sample of 200 households.

Each of the 24 choice sets was evaluated using five independent samples yielding a total of 120 potential parameter estimates (for each specification). Depending on the choice set, each of these estimated models is used to estimate welfare for one or more types of choice sets, for a total of 180 potential combinations of models and true choice sets. With five draws for each combination of true and sampled choice sets, there are 900 observations for the meta analysis.

4. Digression on the Efficient Sample Size

Before proceeding to the results of this analysis, there is one practical issue with sampling that must be considered. Sampling with very large data sets requires a tradeoff between sampling more observations (i.e., housing “choice occasions”) versus more alternatives from the choice set at the occasion of each housing purchase. That is, for any “budget” of computer memory, one could have a data set consisting of a relatively small number of observations, each including a large sampled choice set, or more observations, each with smaller sampled choice sets. This issue has not been previously addressed in the literature, largely because most applications for recreation or transportation involve surveys and, hence, fairly manageable samples. In contrast, models of housing choice based on public records may involve very large samples. In one of the few discussions of the problem, Ben-Akiva and Lerman (1985) suggest that observations would generally provide more information, but the extent of the tradeoff has never been determined. Accordingly, prior to investigating the economic aspects of defining choice sets for RUM applied to housing, we investigate this issue and offer limited evidence in support of their conjecture.

To gauge the comparative information offered by each type of data, we define four distinct combinations of observations and alternatives (with the same total number of cells), repeatedly draw random samples for each combination, and estimate conditional logit models using those samples. Our evaluation of the tradeoff is based on a comparison of the standard errors for the estimated parameters of interest under each combination, with the goal of identifying the more efficient combination (under the assumption that the model is correctly specified).

We consider models with 40,000 observations each with choice sets of 15 alternatives; 30,000 observations with choice sets of 20 alternatives; 20,000 observations with choice sets of 30 alternatives; and 15,000 observations with choice sets of 40 alternatives. Each possible

combination is sampled from the larger data set 300 times. We estimate a logit model with the specifications described above using a sampled choice set of the entire metro area, the entire budget, and a five-month time window, and calculate the standard errors for the parameters for imputed income and the count of days without an ozone alert. Under the assumption that the models are correctly specified, these standard errors can then be compared across types of data sets as a gauge of the effects of the combinations of observations (or “choice occasions”) and choice alternatives.

Table 2 summarizes our findings. For both the ozone and the income parameters, with both the logarithmic and square root preference specifications, the standard errors for the estimated coefficients are smaller the greater the number of observations relative to the number of alternatives in the choice set. This finding is consistent with Ben-Akiva and Lerman’s (1985) conjecture that observed choices contribute more to enhancing the efficiency of estimation.

5. Results

Based on the findings of this comparison, the samples used for our evaluation of the choice set consist of 40,000 observations, each with 15 alternatives. The number of sampled alternatives at this lower end is consistent with previous discrete choice models of housing.¹⁷ It is also close to the average number of houses “considered” (defined by in-person visits) in a sample of house-seekers in Boston (Newberger 1995).

The two preference specifications are each estimated for each of the five draws from each of 24 alternative choice sets. As an illustration of the results from these 240 models, Table 3 presents one set of estimates each for the logarithmic and square root models using 40,000 observations with 15 alternatives, the full budget choice set, a one-month time window to describe the temporal dimension of the choice set, and a five-county search area. The models yield statistically significant estimates for most parameters. Moreover, the sign of each estimated parameter generally agrees with our *a priori* expectations. The only insignificant parameter estimates that arose were for the cases of the qualitative variable for a fireplace and our measure of public safety.

¹⁷ Quigley (1985) uses choice sets consisting of five alternatives in estimation, and Chattopadhyay (2000) uses choice sets of 12 alternatives. Palmquist and Israngkura (1999) use choice sets of 40 alternatives.

It is important to note that some choice sets do lead to implausible results. In particular, when the choice sets were limited to a 44 percent budget share, the coefficient on income was estimated to be negative. The same result was found for the threshold of 52 percent for the case of the square root model. This finding is similar to that of Parsons and Hauber (1998), who conducted a sensitivity analysis over spatially delineated choice boundaries using models of recreation trips, where the price was a function of travel distance. They estimated much lower (in absolute magnitude) price coefficients in the models with the smallest geographic boundaries (i.e., the lowest price threshold). In our case, the negative marginal utility of income implies that welfare measures derived from the models will be inconsistent with the maintained theory underlying the choice model. As a result, these choice set combinations were dropped. Of the original 180 possible combinations of choice set boundaries at the estimation and welfare stages, dropping these leaves 90 and 30 combinations for the logarithmic and square root cases respectively.

We summarize the results from these models using separate meta analyses for each preference specification. In each case, the dependent variable is a numerical estimate of the WTP for nonmarginal changes in ozone. The regressors are indicator variables for the boundaries in each dimension, separately taken at the estimation stage (sampling) and the welfare stage (the “true” choice set). The observations in the meta analysis are mean WTP (averaged over simulated households) for each estimated model. Again, with 90 remaining boundary combinations at the two stages for the log case (respectively, 30 for the square root case), each sampled five times, there are 450 observations available (respectively, 150).

Table 4 presents the results from the meta analysis. The estimated coefficients are in most cases highly significant, but the fit is good for the log case only, with an R^2 of 0.74 compared to 0.12 in the square root case. Note that predicted average welfare estimates for the hypothesized large air quality improvement can be calculated by summing any valid combination of choice set indicators. For the logarithmic specification, estimated expected values range from \$216 to \$4,051; for the square root specification, they range from \$76 to \$599.

We use these meta models to draw inferences about the influence of the attributes of the choice sets for our WTP estimates. Consider first the temporal boundary based on the likely period houses are on the market and on the period households search. At the estimation stage, the table indicates that there is no systematic pattern in the welfare measures. In the case of the logarithmic specification, welfare measures are higher in the three-month window relative to the one-month window, but smaller again in the five- and seven-month windows. These rankings

are almost reversed in the square root specification. Moreover, the effects are small though statistically significant.

Similarly, at the welfare stage, there is no discernable pattern in the effect of the time window in the complete choice set, with all the estimates being statistically insignificant. These findings are consistent with the fact that the time windows add potential alternatives that are uncorrelated with the economic variables and hence with utility. One might argue that at the estimation stage, a smaller time window would be more appropriate since houses in this range are more likely to be in the true choice set. Using this narrower window would be consistent with importance sampling (see Train et al. 1987).

A clearer pattern emerges with respect to the budget boundary. As mentioned previously, use of a tighter boundary of 44 percent of income leads to a negative sign on income in both specifications. The same is true with the 52 percent boundary in the square root specification, so those indicators were dropped from the meta analysis. Furthermore, for the logarithmic specification, the 100-percent boundary at the stage of estimation lowers welfare estimates relative to the 52-percent boundary. This finding is consistent with the estimated marginal utility of money (relative to ozone) being higher under the 100-percent boundary. As suggested earlier, this may be because identification of the income parameter relies on the most expensive houses.

In the case of the logarithmic specification, the 100-percent income boundary at the *welfare* stage also lowers estimated values. This may be because when households are allowed to select expensive houses, they move along the nonlinear utility function to a higher marginal utility of income, thereby decreasing their marginal willingness to pay for ozone improvements (at the selected alternative). In any event, the differences caused by the budget boundaries account for most of the variation in the data. This is why the R^2 is larger for the logarithmic model, which includes budget dummies, than for the square root model.¹⁸

With respect to the spatial boundaries, no pattern is discernable at the estimation stage. For the logarithmic specification, welfare values are lower for models estimated with choice sets that include all counties; for the square root specification, the reverse is true. On the other hand, at the welfare stage, both specifications imply that including the entire metropolitan area in the true choice set increases welfare values.

¹⁸ Estimating the meta analysis for the square root model with the negative value-of-money models increases the R^2 to 0.82.

To further verify these findings, mean *marginal* values for ozone improvements (based only on the estimated parameters and not on true choice sets) were regressed on the five indicator variables at the estimation stage. Marginal values are based only on estimated marginal rates of substitution, which are functions of the preference parameters but not the additive error terms, or, hence, the true choice set. Thus, this meta analysis contains variables for the estimation stage only. The results are summarized in Table 5. All coefficients have the same signs as the corresponding coefficients on the estimation-stage indicator variables in Table 4, thus confirming the findings. As before, estimated marginal values for a one-day decrement in the number of ozone violations can be determined for each choice set from the parameters in the table. They range from \$20 to \$442 for the logarithmic specification and from \$7 to \$53 for the square root specification.

6. The Effects of Choice Sets in Other Applications

Although not explicitly studied in this way, the importance of choice sets in discrete choice modeling has been recognized in a variety of empirical contexts. Nevo (2003) explicitly values a change in the available choice set of breakfast cereals. Recent work by Berry, Levinsohn, and Pakes (2003) on automobile choices includes sports utility vehicles (SUVs), which were omitted from the choice set in their earlier work (1995).

Outdoor recreation has been one of the most widespread applications and one where analysts have explicitly recognized the importance of judgments regarding choice sets. Analysts have considered a number of issues, including the level of aggregation in choice alternative; the factors influencing composition of choice sets, such as information, distance, and activity; and the tailoring of choice sets based on the objective of the analysis (see Haab and Hicks 1997, Parsons et al. 2000). In work closest to the issues explored here, Parsons and Hauber (1998) test the sensitivity of results to successively expanding the scope of households' choice sets in a single dimension: the maximum travel distance to recreation sites. They find that, at first, increasing the maximum allowable travel distance increases the estimated price coefficient on travel distance and decreases welfare measures. Then, at the one hour to one and one-half hour mark, welfare measures flatten and become robust to further widening of the choice set.

Unfortunately, it is not clear whether Parson and Hauber's results reflect sensitivity to modeling judgments about sampling at the estimation stage or about the true choice set required for welfare computations. The confusion arises because they estimate models using lakes that are sampled from within a given boundary and calculate welfare values for an improvement in

water quality under the assumption that the boundary for estimation is also the boundary for the true choice set. By computing welfare estimates for all potentially true boundaries for which the estimation boundaries are theoretically valid, we have distinguished between these effects in our analysis.

Parson and Hauber's findings could be consistent with the importance of distance at either stage. At the estimation stage, just as we needed sufficient variation in housing prices to identify the income parameter in our model, they may need greater allowable travel distances to identify the travel-cost (price) parameter. At some point, additional variation is no longer required and their estimates flatten. In general, alternatives with a higher probability of being chosen are likely to provide more information for the estimator. Parson and Hauber's results are consistent with this observation, inasmuch as they find that adding the most distant lakes to the choice set (with low probabilities of being chosen) does not substantively affect the model estimates.¹⁹

On the other hand, Parsons and Hauber's results are also consistent with the importance of the choice set boundary when computing welfare values, if the sites cleaned up in their welfare scenarios are correlated with average distance. For example, if the dirtiest sites were near an urban city where most people live, then values for cleaning up those sites might decline as the assumed true choice set is widened so as to include more clean substitutes.²⁰

Again, our practice of computing welfare estimates on all the larger choice sets consistent with each estimated model, and the related use of separate indicators in our meta analysis for boundaries at each stage, is an attempt to address precisely these issues. We have also expanded the dimensions in which to investigate the extent of choice set boundaries to include time and

¹⁹ For this reason, Train (2003) suggests using importance sampling that departs from the uniform conditioning property.

²⁰ Further complicating the matter is the fact that Parson and Hauber's strategy actually violates McFadden's positive conditioning property. They include all observed trips in their data, regardless of the definition of the true choice set. Thus, for smaller definitions of the boundary, they include observed trips outside the boundary, but match them only with nonchosen alternatives from within the boundary. This violates positive conditioning since, if another site inside the boundary had been chosen, the probability of sampling this choice set would be zero. It would seem that this might explain their finding that the coefficient on travel cost at first rises as they expand the boundary: estimation with small boundaries includes choice occasions where households are observed to travel great distances relative to a sampled set of alternatives that are all nearby, thus biasing downward the coefficient. However, they note that in fact there was little substantive difference when these observations were simply dropped (see their note 10, p. 40).

money as well as space. With these added dimensions, the meta analysis becomes a more important tool for synthesizing the many results.

7. Summary and Conclusions

There are two separate issues raised by this research. The first is general and considers the way we report the findings from empirical research when a model or testing strategy requires analyst judgment. We find it hard to imagine cases where such judgments are not a key part of the process of developing estimates of the parameters of interest or the test statistics. Conventional practice has encouraged authors to report results that “make the best case” for their proposed model’s estimates or their test conclusion. Even when authors would like to present more sensitivity analyses, they may not be allocated more scarce journal space. This outcome limits the sharing of modeling experience among analysts, and impinges on the application of empirical work to business or public policy.

By expanding the application of meta analysis, it is possible to modify this practice and enhance our understanding of the importance of the “details” underlying each empirical analysis. Meta summaries of the impacts of model judgments on a central variable of interest, whether parameter estimate or test result, enhance the ability of readers to judge the robustness of conclusions and identify the areas for further research. They can also present voluminous information in a compact form. Of course, it is important to acknowledge that our particular application had a very large sample allowing random sampling of recorded housing sales and a quasi-experimental format for our evaluation of modeling judgments. In many applied situations, samples are smaller and it is not possible to use random samples as a basis of evaluating alternative judgments. In these cases, bootstrapped samples might be considered as a strategy to evaluate alternatives. This would not eliminate the correlation between estimates across models and the need to adjust for the nonspherical errors in estimated meta summary functions.²¹

The second, more specific, contribution of this research is to the growing popularity of the random utility model with applications to public goods. A key underlying assumption of the framework has been the choice set. Welfare estimates are sensitive to modeling judgments about how to sample choice sets for estimation and how to define the true choice set for purposes of

²¹ Of course, this same problem arises when meta samples include more than one set of results from a given sample.

welfare measurement, at least in some dimensions. Our results suggest it is less sensitive, at least in a systematic way, to dimensions such as time that are not correlated with the attributes of the model. In these cases, a “conservative” approach that includes the most likely alternatives in the choice set may be appropriate. Our findings also suggest that welfare measures are more sensitive to dimensions such as income/price that are key attributes of the alternatives.

References

- Banzhaf, H. Spencer. 2002. Quality Adjustment for Spatially-Delineated Public Goods: Theory and Application to Cost-of-Living Indices in LA. Discussion Paper 02-10. Washington, DC: Resources for the Future.
- Bayer, Patrick. 2000. Exploring Differences in the Demand for School Quality: An Empirical Analysis of School Choice in California. Mimeo. New Haven, CT: Yale University.
- Ben-Akiva, Moshe, and Bruno Boccara. 1995. Discrete Choice Models with Latent Choice Sets. *International Journal of Research in Marketing* 12: 9-24.
- Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63: 841-890.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 2003. Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*, forthcoming.
- Brock, William A., and Steven N. Durlauf. 2001. Discrete Choice with Social Interactions. *Review of Economic Studies* 68: 235-260.
- Brock, William A., and Steven N. Durlauf. 2002. A Multinomial Choice Model of Neighborhood Effects. *American Economic Review, Papers and Proceedings* 92: 298-303.
- California Association of Realtors. 2000. Median Time on Market for Single-Family Homes, Condos in California Falls in January 1997. <http://www.car.org/newsstand/news/march97-3.html>.
- Chattopadhyay, Sudip. 2000. The Effectiveness of McFaddens's [sic] Nested Logit Model in Valuing Amenity Improvements. *Regional Science and Urban Economics* 30: 23-43.
- Cropper, Maureen L., Leland 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* 75: 225-232.

- Ellison, Glenn. 2002. The Slowdown in the Economics Publishing Process. *Journal of Political Economy* 110: 947-993.
- Haab, Timothy C., and Robert L. Hicks. 1997. Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand. *Journal of Environmental Economics and Management* 34: 127-147.
- Hanemann, W. Michael. 1978. A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement. Unpublished Ph.D. Thesis, Harvard University.
- Herriges, Joseph A., and Catherine Kling. 1999. Nonlinear Income Effects in Random Utility Models. *Review of Economics and Statistics* 81: 62-72.
- Horowitz, Joel L., and Jordan J. Louviere. 1995. What is the Role of Consideration Sets in Choice Modeling? *International Journal of Research in Marketing* 12: 39-54.
- Kaoru, Yoshiaki, and V. Kerry Smith. 1990. 'Black Mayonnaise' and Marine Recreation: Methodological Issues in Valuing Cleanup. RFF Discussion Paper QE91-02. Washington, DC: Resources for the Future.
- McFadden, Daniel. 1978. Modeling the Choice of Residential Location. In *Spatial Interaction Theory and Planning Models*. Edited by Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jörgen W. Weibull. Amsterdam: North-Holland.
- McFadden, Daniel. 1999. Computing Willingness-to-Pay in Random Utility Models. In *Theory and Econometrics: Essays in Honour of John S. Chipman*. Edited by J. Moore, R. Riezman, and J. Melvin. London: Routledge.
- Nevo, Aviv. 2003. New Products, Quality Changes and Welfare Measures Computed from Estimated Demand Systems. *The Review of Economics and Statistics* 85(2): 266-275.
- Newberger, H. 1995. Source of Differences in Information Used by Black and White Housing Seekers. *Urban Studies* 32: 445-470.
- Palmquist, Raymond B. 2003. Property Value Models. In *Handbook of Environmental Economics*. Edited by Karl-Göran Mäler and Jeffrey Vincent, forthcoming.
- Palmquist, Raymond B., and Adis Israngkura. 1999. Valuing Air Quality with Hedonic and Discrete Choice Models. *American Journal of Agricultural Economics* 81: 1128-1133.
- Parsons, George R., and A. Brett Hauber. 1998. Spatial Boundaries and Choice Set Definition in a Random Utility Model of Recreation Demand. *Land Economics* 74: 32-48.

- Parsons, George R., and Mary Jo Kealy. 1992. Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation. *Land Economics* 68: 93-106.
- Parsons, George R., Andrew J. Plantinga, and Kevin J. Boyle. 2000. Narrow Choice Sets in a Random Utility Model of Recreation Demand. *Land Economics* 76: 86-99.
- Peters, T., W.L. Adamowicz, and P.C. Boxall. 1995. Influence of Choice Set Considerations in Modeling the Benefits from Improved Water Quality. *Water Resources Research* 31: 1781-1787.
- Poterba, James M. 1992. Taxation and Housing: Old Questions, New Answers. *American Economic Review Papers and Proceedings* 88(2): 237-242.
- Quigley, John M. 1985. Consumer Choice of Dwelling, Neighborhood, and Public Services. *Regional Science and Urban Economics* 15(1): 1-23.
- Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation*. Cambridge, MA: Cambridge University Press.
- Train, Kenneth E., Daniel McFadden, and Moshe Ben-Akiva. 1987. The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choice. *Rand Journal of Economics* 18: 109-123.

Table 1. Number of Observations and Means of Housing Variables by County, 1989-1994.

Variable	Los Angeles	Orange	Riverside	San Bernardino	Ventura
Price (nominal \$)	234,674	262,891	138,025	150,236	235,151
Lot size (acres)	0.18	0.16	0.24	0.21	0.22
Building size (sq. ft)	1,568	1,766	1,629	1,619	1,831
Bathrooms	1.92	2.17	2.07	2.11	2.24
Bedrooms	3.03	3.33	3.26	3.29	3.47
Fireplace (0/1)	0.54	0.20	0.84	0.80	0.80
Swim Pool (0/1)	0.16	0.13	0.12	0.13	0.15
Age of house	39.2	25.7	10.9	17.3	19.1
Days w/o ozone alert	334.5	356.4	312.4	286.8	359.9
Teachers per student	0.041	0.041	0.040	0.040	0.040
Test score*	4.79	5.68	4.80	4.82	5.04
Public safety**	2,374	2,443	2,282	2,328	2,604
1 Mile of coast (0/1)	0.03	0.06	0	0	0.04
Median income (\$)	47,027	59,166	40,444	44,054	51,325
Number of Observations	144,731	71,147	43,588	37,686	22,489

* The mean math score for 10th grade tests at the school district level.

** Public safety is measured as 3,000 minus the crime rate per 10,000 residents.

Table 2. Average Asymptotic Standard Error of Ozone and Income Parameters under Four Observation/Choice Set Combinations (300 draws)

Parameter	Combination (Observations/Size of Choice Set)			
	(40,000/15)	(30,000/20)	(20,000/30)	(15,000/40)
Ozone, logarithmic	0.0700	0.0854	0.1033	0.1236
Income, logarithmic	0.0385	0.0431	0.0555	0.0627
Ozone, square root	0.0078	0.0095	0.0116	0.0138
Income, square root	3.18e-4	3.61e-4	4.83e-4	4.96e-4

Table 3. One Sample's Results for RUM with the Choice Set Defined by: 1-Month Time Window, 5-County Search Area, and 100% Budget

Independent Variable	Specification for Indirect Utility Function	
	Logarithmic	Square Root
Orange Co.	-0.1456 (0.0193)	-0.1077 (0.0194)
San Bernardino Co.	-0.2425 (0.0218)	-0.1488 (0.0221)
Riverside Co.	-0.3301 (0.0221)	-0.2008 (0.0225)
Ventura Co.	-0.1273 0.0235	-0.0824 (0.0238)
Coast	0.4531 (0.0338)	0.3354 (0.0338)
Air Quality	0.4075 (0.0719)	0.0294 (0.0081)
Teacher-Student Ratio	0.7835 (0.1217)	5.8814 (1.2076)
Test Score (math)	0.6698 (0.0651)	0.4085 (0.0581)
Public Safety (crime rate)	-0.1372 (0.0881)	-0.0071 (0.0039)
Bedrooms	-0.1903 (0.0257)	-0.1698 (0.0304)
Bathrooms	0.0347 (0.0237)	0.0563 (0.0371)
Building Size (sq. ft)	0.6247 (0.0296)	0.0236 (0.0015)
Lot Size (sq. ft)	0.1282 (0.0126)	0.0016 (0.0002)
Fireplace	0.0085 (0.0125)	0.0094 (0.0123)
Pool	0.061 (0.0152)	0.0401 (0.0151)
Age	-0.0134 (0.0066)	0.0025 (0.0039)
Net Income	1.3547 (0.0311)	0.0083 (0.0003)

Standard errors shown in parentheses.

Table 4: Meta Analysis of Welfare Estimates as a Function of Choice Set Boundaries for Estimation and Welfare Computation

	Logarithmic Model	Square Root Model
Constant	2522.34*** (101.01)	330.96 *** (38.37)
Estimation: 3 months	123.89* (69.70)	-13.71 (28.60)
Estimation: 5 months	-217.56*** (81.73)	84.41 ** (33.54)
Estimation: 7 months	-129.14 (107.41)	-31.20 (44.07)
Estimation: 100% of income	-1720.14*** (66.12)	--
Estimation: all counties	-799.67*** (66.12)	34.18 (27.13)
Welfare: 3 months	72.92 (110.21)	-5.59 (45.22)
Welfare: 5 months	2.86 (107.42)	-8.05 (44.07)
Welfare: 7 months	34.37 (107.42)	-3.79 (44.07)
Welfare: 100% of income	-251.04*** (66.12)	--
Welfare: all counties	256.16*** (66.12)	40.58 *** (27.13)
R-Squared	0.74	0.12
N	450	150

Standard errors shown in parentheses. Terms in brackets are the estimated standard errors using sample of group means.

*Significant at 10-percent level based on the micro sample. **Significant at 5-percent level. ***Significant at 1-percent level.

Table 5. Meta Analysis of Marginal Welfare Estimates as a Function of Choice Set Boundaries at the Estimation Stage

	Logarithmic Model	Square root Model
Constant	245.05*** (4.82)	31.65 *** (1.67)
Estimation: 3 months	17.89*** (6.23)	-0.96 ** (2.33)
Estimation: 5 months	-19.78*** (7.06)	6.77 (2.64)
Estimation: 7 months	-3.94 (9.12)	-3.32 (3.41)
Estimation: 100% of income	-179.50*** (5.47)	
Estimation: all counties	-73.80*** (0.41)	3.48 * (2.05)
R-Squared	0.74	0.09
N	450	150

Standard errors shown in parentheses. Terms in brackets are the estimated standard errors using sample of group means.

*Significant at 10-percent level based on the micro sample. **Significant at 5-percent level. ***Significant at 1-percent level.