Division of Agricultural Sciences
UNIVERSITY OF CALIFORNIA

Working Paper No. 156

A QUALITATIVE-QUANTITATIVE MODEL OF CONSUMER CHOICE WITH AN APPLICATION TO RECREATION DEMAND

by

W. Michael Hanemann

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
September 1980

Division of Agricultural Sciences
UNIVERSITY OF CALIFORNIA

A QUALITATIVE-QUANTITATIVE MODEL OF CONSUMER CHOICE WITH AN APPLICATION TO RECREATION DEMAND

W. Michael Hanemann

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
September 1980

A QUALITATIVE-QUANTITATIVE MODEL OF CONSUMER CHOICE WITH AN APPLICATION TO RECREATION DEMAND

By W. Michael Hanemann[1]

1. INTRODUCTION

The development of the multinomial logit (MNL) model and, subsequently, of other types of the quantal choice model has significantly improved the set of tools available for the empirical analysis of consumer demand.[2] With these models, it becomes possible to analyze purely qualitative data on consumer choices—e.g., the choice among alternative brands of a commodity. In some cases, however, it would be useful to have a model which permits both qualitative and quantitative choices—e.g., a consumer selects one brand of a commodity and has to decide both which brand and how much of the brand to buy. This type of choice has appeared in recent theoretical work by Lancaster [15] and Novshek and Sonnenschein [20]. My concern here is with the empirical formulation and estimation of such models. I shall describe two general utility models which are consistent with this type of qualitative-quantitative choice and which can be estimated empirically, and I shall apply one of these models to micro data on the visitation of recreation sites in the Boston area.

The existing literature contains several statistical models for the analysis of qualitative-quantitative or "mixed continuous/discrete" dependent variables, starting with the Tobit model and its subsequent generalization by Heckman and others (see, for example, [12] and [17]). An important practical feature of the latter models is that their estimation can often be decomposed into two stages, one based on the qualitative choice and the other based on the

quantitative choice. However, in these statistical models the qualitative and quantitative choices are not integrated in the sense that they do not both flow from the same underlying utility (or production) function. Duncan's recent work [6], which is set in a production context, partially rectifies this. In his example, the qualitative element is the choice of a location at which to produce, and the quantitative element is the level of inputs and output at that location; both choices are based on the same underlying profit function. One question is still unresolved: how does the qualitative choice arise or why is the economic agent led to a corner solution for his maximization problem In some cases the corner solution is logically necessary because the qualitative choices are mutually inconsistent (e.g., to join or not join a union). In other cases this is not so. Why could a firm not choose to produce at both locations or a consumer choose to buy both brands of a good? If a corner solution occurs in these cases, it must be due to the special nature of the underlying production technology or consumer preferences. This feature is built into the utility models which are presented below; unlike Duncan's production model, they are explicitly structured in such a way as to preclude interior solutions.

The empirical analysis of recreation demand has been the subject of two previous papers in this journal: by Burt and Brewer [3] and by Cicchetti, Fisher, and Smith [4]. Both papers estimate a set of linear ordinary demand functions for alternative recreation sites. This type of demand model can be criticized because it does not explicitly incorporate quality variables which may differentiate the sites, and it is not directly consistent with a utility maximization model.[3] In both cases, the fitted demand functions are used to construct an estimate of the change in consumer's surplus arising from changes

in the prices of recreation sites due to the opening of a new site. However, since the demand functions are not consistent with a utility model, this estimate is not exact. Moreover, since the demand model does not explicitly incorporate site quality, it cannot be used to construct welfare measures for changes in quality rather than price. The demand model estimated in this paper will remedy both of these defects.

The organization of the paper is as follows. Section 2 describes two general qualitative-quantitative choice models; in both models the estimation can be conducted in two stages, one using the qualitative choice data and the other using the quantitative choice data. Section 3 describes in detail the specification and estimation of a version of one of these models, based on the work of Blackburn [1, 2], and explains its relation to the MNL model. It turns out that, when slightly reparameterized, Blackburn's model subsumes the MNL model and extends it in a natural manner to qualitative-quantitative choices. In Section 4 the model is applied to the recreation demand data, and the fitted model is employed in Section 5 to calculate welfare measures for hypothetical price and quality changes. The resulting estimates of consumer's surplus per visitor day are found to be significantly smaller than those currently used for recreation planning. In both sections the model is contrasted with an *ad hoc* demand model, not strictly consistent with a utility maximization hypothesis, which is a simplified version of the statistical model in [17]. The latter model is found to fit the data somewhat less well than the utility-theoretic demand model. The conclusions are summarized in Section 6.

2. GENERAL STRUCTURE OF QUALITATIVE-QUANTITATIVE CHOICE MODELS

In this section I shall discuss how one might formulate qualitative quantitative choice models which can be estimated empirically. In these

models I assume that the qualitative choice is restricted to the selection of only one of the differentiated goods available. This is a special case of a more general corner solution model in which the consumers may select <u>any</u> subset of the differentiated goods--not necessarily one of them and not necessarily all of them. The general corner solution problem is considerably more difficult to formulate empirically. In [10, pp. 153-161] I have developed a demand model for this case, but it can be implemented only when the number of brands is very small; this will not be discussed here. The general set-up and notation are as follows. A consumer chooses among N different brands of a commodity. He has to decide which brand to select--say, brand i--how much of it to consume, x_i, and how much to spend on other goods. For simplicity, I assume that the other goods are not differentiated and can be aggregated into a single commodity denoted by z. The prices of the differentiated goods are $p = (p_1, \ldots, p_N)$, the price of the composite commodity is p_z, and the consumer's income is y. The arguments of the consumer's utility function are $x = (x_1, \ldots, x_N)$, z, and $b = (b_{11}, \ldots, b_{NK})$, where b_{jk} is the amount of the *k*th quality characteristic associated with a unit of consumption of brand j. Let $b_j = (b_{j1}, \ldots, b_{jK})$ be the vector of characteristics (attributes) of good j. I assume that they enter the utility function through aggregator functions, $\phi_j = \phi_j(b_j)$. Thus, the utility function may be written compactly as $u(x, \phi, z)$. The choice variables are (x, z), and the utility function is quasi concave in these variables; the consumer maximizes it subject to the budget constraint $y = \Sigma p_j x_j + p_z z$.

In order to provide a statistical framework for estimating the qualitative-quantitative ordinary demand functions which arise as a corner solution to

this utility maximization problem, it is necessary to introduce a stochastic element into the utility function. In the models to be presented below, this will be done by making the aggregators random functions of their arguments, $\tilde{\phi}_j = \tilde{\phi}_j(b_j)$.[4] This can be rationalized in three ways, which have different implications for the formulations of the aggregator functions: (i) *Varying Preferences*--each individual has specific tastes, but tastes vary across consumers. This could be represented by

$$\tilde{\phi}_j = \Sigma\tilde{\gamma}_k \, b_{jk} \tag{1}$$

in which the weights placed on different characteristics vary among consumers or by

$$\tilde{\phi}_j = - \Sigma\gamma_k \, (\tilde{\alpha}_k - b_{jk})^2 \tag{2}$$

which implies that, for every attribute, there is an "ideal" level of the attribute for each individual, $\tilde{\alpha}_k$, which varies among individuals; consumers seek to minimize the weighted distance between actual and ideal attribute levels. (ii) *Pure Probabilistic Behavior*--tastes vary randomly, not only across individuals but also, for a given individual, across alternatives. (iii) *Unobserved Attributes of Alternatives*--the presence of unmeasured attributes of alternatives induces an element of randomness in observed demands. In practice, cases (ii) and (iii) are usually indistinguishable and are modeled in the same way, e.g., by writing

$$\tilde{\phi}_j = \gamma_j(b_j) + \tilde{\varepsilon}_j. \tag{3}$$

Combinations of the three cases are also possible. In the quantal choice literature, Hausman and Wise [11] present examples of (1) and (3), based on

the multivariate normal distribution, and McFadden [18, 19] works with (3) and the extreme value or generalized extreme value distributions; (2) appears in [20] with the uniform distribution and, implicitly, in [15]. Whatever the specification of the random component in the $\tilde{\phi}_j$'s, it induces a cumulative joint distribution function denoted below by $F_{\phi}(t_1, \ldots, t_N)$.

Assuming that the form of utility function $\tilde{u} = u(x, \tilde{\phi}, z)$ is such as to insure a corner solution, one may proceed as follows. Suppose the consumer selects brand i. Set $x_j = 0$, $j \neq i$, and maximize utility subject to the budget constraint with respect to (x_i, z). This yields the ordinary demand functions for x_i and z conditional on the choice of brand i. Substitution into the utility function yields the indirect utility function conditional on the choice of brand i, $\tilde{u}_i = v_i(p_i, p_z, \tilde{\phi}_i, z)$. The brand chosen is that which solves: $\max_j \{\tilde{u}_1, \ldots, \tilde{u}_N\}$. Denote the value of the maximized indirect utility function by $\tilde{u}^*$. Let $F_u(u_1, \ldots, u_N)$ be the joint cumulative distribution function (c.d.f.) of the $\tilde{u}_j$'s induced by $F_{\phi}(\cdot)$, and let $F_u^i(u_1, \ldots, u_N)$ be the partial derivative with respect to the ith argument. In general, these are functions of (p, b, p_z, y). $F_u^i(u, \ldots, u) \equiv \Pr\{\tilde{u}_i = u, \tilde{u}_j \leq u \; j \neq i\}$ may be considered as a function of a single variable, u. The qualitative choice probabilities are given by

$$(4) \qquad p^i \equiv \Pr\{i \text{ chosen}\} = \int_{-\infty}^{\infty} F_u^i(u, \ldots, u)\, du.$$

Note that $F_u(u, \ldots, u)$, regarded as a function of a single variable, is the c.d.f. of $\tilde{u}^*$. Then $F_u'(u, \ldots, u)$, the derivative of this function with respect to the single argument, is the p.d.f. of $\tilde{u}^*$. Thus,

$$\bar{u}^* = \int_{-\infty}^{\infty} u \cdot F_u' (u, \ldots, u)\, du \tag{5}$$

is the expected value of the maximized indirect utility function. To emphasize its dependence on these variables, I will write $\bar{u}^* = \bar{v}^* (p, b, p_z, y)$. This function can be employed to construct monetary measures of the welfare effects of a change in the prices or qualities of the available brands. Suppose that prices and qualities change from $\left(p^0, b^0, p_z^0\right)$ to $\left(p^1, b^1, p_z^1\right)$. The compensating variation measure of the welfare effects of this change is the quantity, C, which satisfies

$$\bar{v}^* \left(p^1, b^1, p_z^1, y - C\right) = \bar{v}^* \left(p^0, b^0, p_z^0, y\right), \tag{6a}$$

while the equivalent variation measure is the quantity, E, which satisfies

$$\bar{v}^* \left(p^1, b^1, p_z^1, y\right) = \bar{v}^* \left(p^0, b^0, p_z^0, y + E\right). \tag{6b}$$

In order to develop the probability formulas for the quantitative choices of the branded and nonbranded goods, one needs to specify further the structure of the utility function. I shall discuss two formulations (each unique up to a monotone-increasing transformation):

$$u (x, \tilde{\phi}, z) = g (\Sigma x_j) + \Sigma \tilde{\phi}_j x_j + hz \qquad g'' < 0, \qquad h > 0 \tag{7}$$

$$u (x, \tilde{\phi}, z) = \bar{u} (\Sigma \tilde{\phi}_j x_j, z). \tag{8}$$

By inspection, it can be seen that, when these functions are maximized subject to the budget constraint, they each have a corner solution in which only one brand is selected.[5]

The utility function (7) is a generalization of the one which underlies Blackburn's models [1, 2]. It may be rewritten as

$$u\ (x,\ \tilde{\phi},\ z) = g\ (\Sigma x_j) + h\left[\frac{y}{p_z} - \ \Sigma\ \tilde{\pi}_j\ x_j\right]$$

where $\tilde{\pi}_j = [p_j/p_z - \tilde{\phi}_j/h]$. The $\tilde{\pi}_j$'s can be regarded as "generalized" or "quality corrected" price ratios. If the $\tilde{\pi}_j$'s are all different, the consumer selects only one brand—that with the lowest $\tilde{\pi}_i$. Conditional on the selection of brand i, the ordinary demand function for the branded good is $\tilde{x}_i = \psi\,(h\tilde{\pi}_i)$ where $\psi \equiv (g')^{-1}$, and the ordinary demand function for the composite commodity is $z = [y - p_i\ \psi\ (h\tilde{\pi}_i)]/p_z$. An important implication is that all of the branded goods have a zero income elasticity of demand, while the composite commodity's income elasticity of demand is equal to the inverse of its budget share. Define $\tilde{T}_j = h\tilde{\pi}_j$. Let $F_T\ (t_1, \ldots, t_N) \equiv \Pr\ \{\tilde{T}_1 > t_1, \ldots, \tilde{T}_N > t_N\}$ be the joint c.d.f. induced by $F_\phi(\cdot)$, and let $F_T^i\ (t_1, \ldots, t_N)$ be the negative of its partial derivative with respect to the *i*th argument, i.e., $F_T^i\ (t_1, \ldots, t_N) \equiv \Pr\ \{\tilde{T}_i = t_i,\ \tilde{T}_j > t_j\ \ j \neq i\}$. The qualitative choice probabilities in this model are given by

$$P^i = \int_{-\infty}^{\infty} F_T^i\ (t, \ldots, t)\ dt. \tag{9}$$

In order to derive the quantitative choice probabilities, continue to regard $F_T^i\ (t, \ldots, t)$ as a function of a single variable and make the change of variable from t to $\psi(t)$.[6] The probability density function (p.d.f.) for x_i, conditional on the choice of brand i, is

$$f_{x_i}(x) \equiv \Pr\{\tilde{x}_i = x,\ \tilde{z} = \frac{y - p_i\, x}{p_z},\ \tilde{x}_j = 0\ j \neq i\}$$

(10)

$$= F_T^i\,[\psi^{-1}(x),\ \ldots,\ \psi^{-1}(x)] \cdot \left| \frac{d\psi^{-1}(x)}{dx} \right|.$$

Note that $\psi^{-1}(\cdot) \equiv g'(\cdot)$. The conditional (and unconditional) expectation of the quantity demanded is

(11) $$E\{x_i\} = \int_0^\infty x \cdot f_{x_i}(x)\, dx.$$

The formulas for the $\tilde{u}_j$'s, and their joint c.d.f. will be given in the next section for a particular specification of $g(\cdot)$.

The utility function (8) is a simplification of the "ruled indifference surface" model presented in [20]. Here the function $\bar{u}\ (t_I,\ t_{II})$ is a standard quasi-concave neoclassical utility function of two arguments. Let $\bar{h}^I\ (p_I,\ p_{II},\ y)$, $\bar{h}^{II}\ (p_I,\ p_{II},\ y)$, and $\bar{v}\ (p_I,\ p_{II},\ y)$ be the ordinary demand functions and the indirect utility function associated with $\bar{u}\ (\cdot,\ \cdot)$. Define $\tilde{\pi}_j = p_j/\tilde{\phi}_j$. If the $\tilde{\pi}_j$'s are all different, the maximization of (8) leads to the selection of a single brand--that with the lowest $\tilde{\pi}_i$. Conditional on the selection of brand i, the ordinary demand function for the branded good is $\tilde{x}_i = \bar{h}^I\ (\tilde{\pi}_i,\ p_z,\ y)/\tilde{\phi}_i$, and the ordinary demand function for the composite commodity is $\tilde{z} = \bar{h}^{II}\ (\tilde{\pi}_i,\ p_z,\ y)$. In general, these demand functions will exhibit nonzero income elasticities of demand, depending on the form of $\bar{v}(\cdot)$. Let $F_\pi\ (\pi_1,\ \ldots,\ \pi_N) \equiv \Pr\{\tilde{\pi}_1 > \pi_1,\ \ldots,\ \tilde{\pi}_N > \pi_N\}$ be the joint c.d.f. induced by $F_\phi(\cdot)$, and let $F_\pi^i\ (\pi_1,\ \ldots,\ \pi_N)$ be the negative of its derivative with respect to the ith argument. The quantitative choice probabilities for

this model are given by (9), with $F_{\pi}^{i}(\cdot)$ substituted for $F_{T}^{i}(\cdot)$. By writing the conditional ordinary demand function for the branded good in the form $\tilde{x}_i = \psi(\tilde{\pi}_i) \equiv \tilde{\pi}_i \cdot \bar{h}^I (\tilde{\pi}_i, p_z, y)/p_i$, one can make the change of variable from π to $\psi(\pi)$ and obtain the p.d.f. for $\tilde{x}_i$ conditional on the choice of that brand. The formula is similar to (10), with $F_{\pi}^{i}(\cdot)$ substituted for $F_{T}^{i}(\cdot)$. Likewise, the formula for the conditional and unconditional expectations of the demand for the branded good is similar to (11). The formula for the joint c.d.f. of the $\tilde{u}_j$'s, $F_u(\cdot)$, is obtained from $F_{\pi}(\cdot)$ by change of variable, based on the relation $\tilde{u}_i = \bar{v}(\tilde{\pi}_i, p_z, y)$.

In order to implement these models, one needs to specify the form of the random functions $\tilde{\phi}_j(b_j)$ which, in turn, involves picking one of the specifications, (1), (2), or (3). For (7), one also needs to specify the form of $g(\cdot)$; and, for (8), one needs to specify the form of $\bar{u}(\cdot)$. Both models have the property that their estimation can be conducted in two stages. In the first stage one applies the data on qualitative choices to the qualitative choice probability formulas, (4) or (9), to obtain estimates of the coefficients of the $\tilde{\phi}_j(b_j)$ functions. This can be accomplished by using existing computer routines for the maximum likelihood or minimum chi-squared estimation of qualitative choice models such as MNL or multinomial probit, depending on the assumed distribution of the $\tilde{\phi}_j$'s. In the second stage, one estimates the remaining coefficients of the utility function by applying maximum likelihood to the quantitative choice probabilities, (10), or, more simply, by least-squares estimation based on the expected quantity-demanded formula, (11). Thus, by virtue of the decomposition, these qualitative-quantitative choice models can be estimated with existing software.[7] The

estimation procedure is illustrated below for a particular specification of the utility model (7). It is also being applied to the utility model (8), but this will not be reported here.

3. SPECIFICATION AND ESTIMATION OF BLACKBURN'S MODEL

In this section I describe the estimation of a particular version of the qualitative-quantitative choice model (7) due to Blackburn [1, 2]. I also contrast this model with the MNL qualitative choice model which it subsumes and extends to quantitative choices. Both the Blackburn and MNL models employ specification (3) for the $\tilde{\phi}_j(b_j)$ functions, with the $\tilde{\varepsilon}_j$'s independently distributed according to the extreme value (EV) distribution with parameters (μ, α_j), $\mu > 0$.[8] Hence, one can write $\tilde{\phi}_j = \alpha_j + \gamma_j(b_j) + \tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is EV $(\mu, 0)$. It will be shown below that the MNL model is obtained by omitting the $g(\cdot)$ function from (8). In the Blackburn model, $g(s) = s\,[1 + \ln \theta - \ln s]$, where θ is a positive constant.[9] At this point, one can set $p_z = 1$. Thus, Blackburn's utility function is:[10]

$$u\,(x, \tilde{\phi}, z) = \Sigma x_j\,[1 + \ln \theta - \ln \Sigma x_j] + hy - \Sigma \tilde{T}_j\, x_j \tag{12a}$$

where

$$\tilde{T}_j \equiv h\tilde{\pi}_j = hp_j - \gamma_j(b_j) - \alpha_j - \tilde{\varepsilon}. \tag{12b}$$

The unknowns in the model are θ, h, the coefficients of the N functions $\gamma_j(b_j)$, μ, and $\alpha_1, \ldots, \alpha_N$. I will define the following terms for future use. Let $\kappa_j = \alpha_j + \gamma_j(b_j) - hp_j$; then, $\tilde{T}_j = -\kappa_j - \tilde{\varepsilon}$. Let $\bar{\kappa}_j = \kappa_j/\mu$, $\lambda_j = e^{\kappa_j}$, $\bar{\lambda}_j = e^{\bar{\kappa}_j}$, $\lambda = \Sigma\lambda_j$, and $\bar{\lambda} = \Sigma\lambda_j$.

Given that $\tilde{\varepsilon}$ is EV $(\mu, 0)$, the joint c.d.f., $F_T(t_1, \ldots, t_N) \equiv \Pr\{\tilde{T}_1 > t_1, \ldots, \tilde{T}_N > t_N\}$, takes the form

$$F_T(t_1, \ldots, t_N) = \exp\left[-\Sigma e^{(\kappa_j + t_j)/\mu}\right].$$

Thus,

$$F_T^i(t, \ldots, t) = \frac{\bar{\lambda}_i}{\mu} e^{t/\mu} \exp\left[-\bar{\lambda} e^{t/\mu}\right].$$

Accordingly, the qualitative choice probabilities, $P^i \equiv \Pr\{i \text{ chosen}\}$, are given by

$$\begin{aligned} P^i &= \int_{-\infty}^{\infty} \frac{\bar{\lambda}_i}{\mu} e^{t/\mu} \exp\left[-\bar{\lambda} e^{t/\mu}\right] dt \\ &= \int_0^{\infty} \bar{\lambda}_i \, e^{-\bar{\lambda}\omega} d\omega \\ &= \frac{\bar{\lambda}_i}{\bar{\lambda}} \\ &\equiv \frac{\exp\left[\frac{\alpha_i}{\mu} + \frac{\gamma_i(b_i)}{\mu} - \frac{hp_i}{\mu}\right]}{\Sigma \exp\left[\frac{\alpha_j}{\mu} + \frac{\gamma_j(b_j)}{\mu} - \frac{hp_j}{\mu}\right]}. \end{aligned} \tag{13}$$

To obtain the quantitative choice probabilities, note that the conditional ordinary demand function for the branded good is $\psi(\tilde{T}_i) = \theta e^{-\tilde{T}_i}$. Thus, $\psi^{-1}(x) = \ln \theta - \ln x$. Accordingly, the formula for the p.d.f. of $\tilde{x}_i$, conditional on the choice of that brand, corresponding to (10), is

$$f_{x_i}(x) = \frac{\bar{\lambda}_i}{\mu} x^{-(\mu+1)/\mu} \theta^{1/\mu} \exp\left[-\bar{\lambda}\theta^{1/\mu} x^{-1/\mu}\right]. \tag{14}$$

Hence, the expected quantity demanded is

$$E\{x_i\} = \frac{\bar{\lambda}_i\ \theta^{1/\mu}}{\mu} \int_0^\infty x^{-1/\mu} \exp\left[-\bar{\lambda}\theta^{1/\mu}\ x^{-1/\mu}\right] dx$$

$$= \bar{\lambda}_i\ \theta^{1/\mu} \left[\bar{\lambda}\theta^{1/\mu}\right]^{(\mu-1)} \Gamma(1-\mu) \tag{15}$$

$$= \theta\,\Gamma(1-\mu)\ \bar{\lambda}_i\ \bar{\lambda}^{(\mu-1)}$$

where it is required, in order for the integral to converge, that $\mu < 1$. Since the EV distribution implies that $\mu > 0$, the overall requirement is that $0 < \mu < 1$.[11]

To motivate these sign restrictions, consider the following qualitative analysis of the demand formulas (13) and (15). Suppose there is an increase in the attractiveness of brand 1 alone--p_1 falls and/or b_1 rises. This raises $\bar{\lambda}_1$ and $\bar{\lambda}$, but not $\bar{\lambda}_j$, $j \geq 2$. By (13), this raises P^1 and lowers P^j, $j \geq 2$. Furthermore, by (15), since $\mu > 0$, it raises the expected consumption of brand 1, $E\{x_1\}$; and, since $\mu < 1$, it lowers the expected consumption of other brands, $E\{x_j\}$ $j \geq 2$. However, the fall in the consumption of the other brands is more than offset by the increase in the consumption of brand 1 since the expected total consumption of all brands, $\Sigma E\{x_j\} = \theta \cdot \Gamma(1-\mu)\ \lambda^{\mu}$, rises. All of this is intuitively plausible. But there is one possibly undesirable feature of (13) and (15) which also arises in the MNL model and results from the use of the EV distribution. This is the Independence of Irrelevant Alternatives Property: the ratio $E\{x_i\}/E\{x_j\} = P^i/P^j = \bar{\lambda}_i/\bar{\lambda}_j$ is independent of (p_ℓ, b_ℓ), $\ell \neq i, j$.

In order to derive the unconditional expected value of the indirect utility function on which the welfare measures are based, substitute the conditional ordinary demand function into (12) to obtain the conditional indirect utility function, $\tilde{u}_i = hy + \theta e^{-\tilde{T}_i} \equiv hy + \theta \cdot \tilde{Q}_i$, say.[12] Note that $F_{Q_i}(s) \equiv \Pr\{Q_i \leq s\} = \exp[-\bar{\lambda}_i s^{-1/\mu}]$. Then,

$$F_u(u, \ldots, u) = \Pi F_{Q_j}\left[\frac{u - hy}{\theta}\right]$$

$$= \exp\left[-\bar{\lambda}\left(\frac{u - hy}{\theta}\right)^{-1/\mu}\right].$$

Making a change of variable from u to $\omega = (u - hy)/\theta$ and applying (5), one obtains

$$\bar{u}^* = \int_0^\infty \frac{\bar{\lambda}}{\mu}[hy + \theta\omega]\,\omega^{-(1+\mu)/\mu} \exp[-\bar{\lambda}\omega^{-1/\mu}]\,d\omega$$

(16)

$$= hy + \theta\Gamma(1 - \mu)\,\bar{\lambda}^{\mu}.$$

One can now apply the formulas for the compensating and equivalent variations, (6a, 6b). Suppose that the prices and attributes of the set of branded goods change from (p^0, b^0) to (p^1, b^1). Then

(17) $$C = E = \frac{\theta}{h}\,\Gamma(1 - \mu)\left[(\bar{\lambda}^0)^{\mu} - (\bar{\lambda}^1)^{\mu}\right]$$

where

$$\bar{\lambda}^t = \sum_j \exp\left[\frac{\alpha_j + \gamma_j(b_j^t) - hp^t}{\mu}\right] \qquad t = 0, 1.$$

The equality of the compensating and equivalent variations should cause no surprise since it was noted above that the conditional ordinary demand functions for the branded goods implied by the general utility model (7) all have zero income elasticities of demand. Hence, the ordinary and compensated demand curves coincide. This also implies that the welfare measures can be represented as areas under ordinary demand curves. For simplicity, suppose that the price and attributes of only good 1 change from $\left(p_1^0,\ b_1^0\right)$ to $\left(p_1^1,\ b_1^1\right)$. Let $\bar{\lambda}_{(1)} = \bar{\lambda} - \bar{\lambda}_1$. Thus, $\bar{\kappa}_1$ changes from $\bar{\kappa}_1^0$ to $\bar{\kappa}_1^1$, while $\bar{\lambda}_{(1)}$ stays constant. From (17) and (15), one obtains

$$C = E = \frac{\theta}{h}\,\Gamma(1-\mu)\left[\left(e^{\bar{\kappa}_1^0} + \bar{\lambda}_{(1)}\right)^{\mu} - \left(e^{\bar{\kappa}_1^1} + \bar{\lambda}_{(1)}\right)^{\mu}\right]$$

$$(18) \qquad = \int_{\bar{\kappa}_1^1}^{\bar{\kappa}_1^0} \frac{\theta}{h}\,\Gamma(1-\mu)\left[e^{\bar{\kappa}_1} + \bar{\lambda}_{(1)}\right]^{\mu-1} e^{\bar{\kappa}_1}\, d\bar{\kappa}_1$$

$$= \frac{1}{h}\int_{\bar{\kappa}_1^1}^{\bar{\kappa}_1^0} E\,\{x_1(\bar{\kappa}_1)\}\, d\bar{\kappa}_1 .$$

At this point, it is convenient to contrast the Blackburn model with the MNL qualitative choice model. The essential difference is the omission of $g(\cdot)$ from (12) since it influences only the quantitative choices. In addition, one restricts the x_j's so that they take only the values 0 or 1. Accordingly, the utility function underlying the MNL model can be represented as

$$(19) \qquad u(x, \tilde{\phi}, z) = hy - \Sigma\tilde{T}_j x_j, \qquad \tilde{T}_j = -\kappa_j - \tilde{\varepsilon}, \qquad x_j = 0, 1.$$

The reason for retaining the term (hy) in (19) is to provide a basis for welfare measures of price and quality change, as explained below. The maximization of (19) with $\tilde{\varepsilon} \sim EV(\mu, 0)$ leads to the qualitative choice probabilities given by (13). The only difference between (13) and the conventional formula for the MNL qualitative choice probabilities, e.g., [5, page 69], is the appearance of μ, which is the scale parameter of the EV distribution. It is clear that, from (13) alone, one cannot obtain a separate estimate of μ. Suppose that $\gamma_j(b_j) = \Sigma_k \gamma_k b_{jk}$, and let $\bar{\alpha}_j = \alpha_j/\mu$, $\bar{\gamma}_k = \gamma_k/\mu$, and $\bar{h} = h/\mu$. If one works only with data on qualitative choices and (13), one can only obtain estimates of the $\bar{\alpha}_j$'s, the $\bar{\gamma}_k$'s, and $\bar{h}$.[13] In this context, it would be natural to set $\mu = 1$, which is how the MNL model is usually presented. As noted above, for the Blackburn model one requires that $\mu < 1$, and one can obtain a separate estimate of μ from the quantitative demand formulas, (14) or (15). Since, in the empirical section of this paper, I will contrast predictions based on the Blackburn model with those derived from a pure MNL model, it is convenient to set down the formulas for the welfare measures associated with the latter model. The development is the same as that leading up to (16) and (17) except that it is based on (19) with $\tilde{\varepsilon} \sim EV(1, 0)$, i.e., it is based on the standard form of the MNL model. For this model, the conditional indirect utility function is $\tilde{u}_i = hy - \tilde{T}_i$. Hence,

$$F_u(u, \ldots, u) = \exp[-\lambda e^{-(u-hy)}],$$

and

$$\bar{u}^* = \int_{-\infty}^{\infty} u\lambda e^{-(u-hy)} \exp[-\lambda e^{-(u-hy)}]\, du$$

(20)

$$= hy + \ln \lambda + 0.522 \ldots \text{ (Euler's constant).}$$

Therefore, for a change from (p^0, b^0) to (p^1, b^1),[14]

$$(21) \quad C^b = E^b = \frac{1}{h} \ln\left(\frac{\lambda^0}{\lambda^1}\right), \qquad \lambda^t = \sum_j \exp\left[\alpha_j + \gamma_j\left(b_j^t\right) - hp_j^t\right] \qquad t = 0, 1.$$

As was noted above, the MNL model plays a role in the two-stage estimation of the Blackburn model. For simplicity, continue to assume that $\gamma_j(b_j) = \Sigma\gamma_k\ b_{jk}$. In the first stage, one applies the MNL model to the qualitative choice data, based on (13), which yields estimates of the $\bar{\alpha}_j$, the $\bar{\gamma}_k$'s, and $\bar{h}$. This leaves μ and θ to be estimated in the second stage--the estimate of μ combined with the first-stage estimates yields estimates of the α_j's, the γ_k's, and h. There are several options with respect to the second-stage estimation. (i) It can be based on (14) and the maximum likelihood method or on (15) and the regression method; the latter is considerably simpler and will be followed here. (ii) Assuming the estimation is based on (15), this can be regarded as a formula for the conditional or the unconditional expectation of the quantity demanded. Under the first interpretation, as the expected demand conditional on the selection of that brand, the regression model is based just on the brand selected--there is one observation per consumer, namely, the amount consumed of the brand he actually selects. I shall refer to this as the "partial sample" case. Under the second interpretation, as the unconditional expected demand, the regression model can be applied to all of the brands--there are N observations for each consumer, namely, his consumption levels of all brands (N - 1 which are zero). I shall refer to this as the "full sample" case. (iii) The regression can be based on a variety of probability distributions. For example, as an approximation, one might postulate the lognormal regression model

$$\tilde{x}_i = E\{x_i\} \cdot e^{\tilde{\omega}}$$

or

$$\tilde{x}_i = [\theta \Gamma (1 - \mu)] \bar{\lambda}_i \, \bar{\lambda}^{(\mu-1)} \, e^{\tilde{\omega}} \tag{22}$$

where $\tilde{x}_i$ is the observed consumption of brand i and $\tilde{\omega}$ is independently identical distributed N $(0, \sigma^2)$. The two coefficients to be estimated in (22) are an exponent, corresponding to $(\mu - 1)$, and a constant term, corresponding to $[\theta \Gamma (1 - \mu)]$, from which an estimate of θ can be derived. In the partial sample case, one can take the logarithm of both sides of (22) and apply OLS.[15] In the full sample case, the logarithmic transform cannot be applied, and one must estimate (22) directly by nonlinear least squares. There is nothing in this regression which constrains the dependent variable to be an integer. An alternative approach, which incorporates this constraint, is to postulate that the observed demands, x_j, are distributed according to the Poisson distribution, with their mean given by (15). For the full sample case, the Poisson model can be estimated by weighted least squares, as set down in [7] and [14]. For the partial sample case, one must employ the truncated Poisson distribution which omits the zero class. This requires maximum likelihood estimation, which is the procedure employed in [2].

4. AN APPLICATION TO RECREATION DEMAND

The purpose of this section is to illustrate the Blackburn model by applying it to data on the demand for water-based recreation sites in the Boston area. The data come from two surveys, both conducted in 1974 and described in more detail in [10]: a survey in their homes of a stratified random sample of 462 households in the Boston SMSA to ascertain which sites they had visited

during the summer of 1974 for swimming and beach recreation activites and the frequency of their visitation; and a survey of 33 major recreation sites in the area to inventory their facilities and collect water samples for chemical analysis. Most of the households visited more than one site, but 106 households visited only one site. Of these households, there were 83 who, among them, visited 20 sites and who form the sample for the present application.[16] These sites, which are listed in the Appendix, include most of the important beaches in the Boston area. Each household is conceived of as selecting one of the 20 sites and making some number of visits to the site over the summer. In the sample, the number of visits by a household ranged from 1 to 100, with a median of about 5.

The utility model is (12). The variables selected for this application are described in Table 1. I shall use five measures of site quality. Two are measures of water quality, COD and PHOS; another measures nonwater aspects of site quality, NUISANCE.[17] The fourth variable, SITE TYPE, is a dummy variable for freshwater as opposed to ocean sites, since there may be district preferences for the two types of site. The fifth variable, MINORITY ATT, is intended to pick up racial segmentation in recreation behavior: at certain sites in the Boston area, an unusually high percentage of visitors are from certain ethnic or racial groups. This phenomenon is handled here by creating a dummy variable which takes the value 1 if the household is from a racial minority group and the site is one of those identified as having a special attractiveness to minority groups, and 0 otherwise. These five variables constitute the components of b_j. Following the custom in recreation demand studies, the price variable, p_j, is taken to be the travel cost, defined as estimated road distance from the household's home to the recreation site

TABLE 1

Variables Used in the Recreation Site Demand Model

Variable	Definition	Type[a]	Mean value
COD	Chemical oxygen demand of water at the site (mg./ℓ)	S	36.45
PHOS	Total phosphorus content of water at the site (mg./ℓ)	S	0.066
SITE TYPE	1 if freshwater, 0 if saltwater site	S	0.15
NUISANCE	1 if the site has heavily urban or noisy setting, 0 otherwise	S	0.45
TRAVEL COST	Cost (in dollars) of traveling by automobile from home to site	H & S	0.72
MINORITY ATT	1 if the household is a minority group and the site is identified as having a special attractiveness to minorities, 0 otherwise	H & S	0.013
EDUC	Highest household education (categories: 1 = elementary school, ..., 4 = some college, ..., 7 = postgraduate education)	H	4.34
SWIMPOOL	1 if household makes frequent use of a private swimming pool, 0 otherwise	H	0.17
AUTO	1 if household owns one or more cars, 0 otherwise	H	0.86
# KIDS	Number of persons under 18 in the household	H	1.13
# ADULTS	Number of persons aged 18 and older in the household	H	2.61

[a] S = varies over sites; H = varies over households.

multiplied by an estimated travel cost of 7 cents per mile (in 1974 prices).[18] In formulating the model, I adopt an entirely "generic" specification for the $\tilde{\pi}_j$'s, setting $\alpha_j = 0$ all j, and writing

$$(23)\quad \gamma_j(b_j) = \gamma_1 \text{ COD} + \gamma_2 \text{ PHOS} + \gamma_3 \text{ SITE TYPE} + \gamma_4 \text{ NUISANCE} + \gamma_5 \text{ MINORITY ATT.}$$

My prior expectation is that γ_1, γ_2, and $\gamma_4 < 0$, $\gamma_5 > 0$, and $\gamma_3 > 0$ or < 0, depending on whether there is a preference for or against freshwater sites. In addition to the formulation in (12), where the price coefficient h is a constant, I investigated the possibility that households' responses to price vary with their education by writing $h = h_0 \cdot e^{-\text{EDUC}}$; this implies that the impedance effect of higher travel costs diminishes as EDUC rises. I also allowed household attributes to influence the frequency of site visitation by making the coefficient θ a function of these variables. Specifically,

$$(24)\quad \theta = \theta_0 \quad \text{EDUC}^{\beta_1} \quad \exp\,[\beta_2 \text{ SWIMPOOL} + \beta_3 \text{ AUTO} + \beta_4 \text{ \# KIDS} + \beta_5 \text{ \# ADULTS}].$$

My prior expectation is that β_1, β_4, and $\beta_5 > 0$, $\beta_2 < 0$, and $\beta_3 >$ or < 0 depending upon whether access to an automobile increases or decreases a household's use of public beaches as opposed to other types of recreation.

As formulated, the model's coefficients are γ_1, γ_2, γ_3, γ_4, γ_5, h (or h_0), μ, θ_0, β_1, β_2, β_3, β_4, and β_5. The first-stage estimation, consisting of the maximum likelihood estimation of (13) applied to the qualitative choice data, yields estimates of $\bar{\gamma}_k = \gamma_k/\mu$, $k = 1, \ldots, 5$; and $\bar{h} = \bar{h}/\mu$ (or $h_0 = h_0/\mu$). The results are presented in Table 2. The signs of these coefficients conform with my expectations, and there appears to

TABLE 2

First-Stage Estimates of the Coefficients of the Blackburn Model

Coefficient	Estimates[a]	
$\bar{\gamma}_1$	- 0.0292 (3.38)	- 0.0286 (3.34)
$\bar{\gamma}_2$	-16.786 (3.16)	-15.52 (3.04)
$\bar{\gamma}_3$	- 2.134 (4.13)	- 2.0163 (3.70)
$\bar{\gamma}_4$	- 0.5313 (1.63)	- 0.3201 (0.99)
$\bar{\gamma}_5$	1.565 (2.41)	1.576 (2.34)
$\bar{h}$	1.369 (6.52)	45.87[b] (5.77)
ln L	-199.68	-199.66
$\hat{R}^2$[c]	0.197	0.197

[a]The number in brackets is the absolute value of the t statistic.

[b]Estimate of $\bar{h}_0$.

[c]Pseudo R^2 statistic [5, p. 123].

be a distinct preference for saltwater sites as against freshwater sites. The specification with a variable h leads to a slight increase in the goodness of fit; thus, there is weak evidence that households with a higher education are deterred less by higher travel costs.

The coefficient estimates in the first column of Table 2 are used to form estimates of $\bar{\lambda}_1, \ldots, \bar{\lambda}_N$, and $\bar{\lambda}$, which are inputs to the second stage of the estimation procedure. Here this is based on the lognormal regression model, (22), since initial experiments with the Poisson model yielded a poorer fit. Three versions of the lognormal model were estimated--the first two based on the partial sample and the third based on the full sample. The results are shown in Table 3. The first column shows the results of applying OLS to the logarithmic transform of (22). The signs of the coefficient conform with my expectations. Frequent usage of a private swimming pool and access to an automobile both reduce visitation of public beaches. Household education and size both increase visitation of public beaches. Household composition is important, too, since the number of adults in the household affects beach visitation more than the number of children.[19] The a priori restriction that $0 < \mu < 1$ was not explicitly imposed in the estimation of the equation; it is satisfied anyway, although the estimate is not statistically significant. As a means of incorporating the restriction explicitly, I reestimated the equation using the Theil-Goldberger [22] mixed estimation procedure for combining prior information on a coefficient with sample information. In this case the prior information is that μ lies in the unit interval. This can be represented as a situation of incomplete extraneous information involving a linear restriction, namely, a prior point estimate of

TABLE 3

Second-Stage Estimates of the Coefficients of the Blackburn Model

	Blackburn Model[a]			Ad Hoc Model
	Estimation Method[b]			
Coefficient	OLS	OLS Mixed Regression	Nonlinear LS	OLS
	1	2	3	4
β_1	0.7245 (1.44)	0.7235 (1.46)	2.0968 (7.70)	0.1740 (0.51)
β_2	-1.0190 (2.31)	-1.0196 (2.33)	-1.0295 (2.13)	-0.6608 (2.18)
β_3	-1.0440 (2.13)	-1.0434 (2.15)	-0.5583 (3.33)	0.3329 (1.00)
β_4	0.1656 (1.32)	0.1655 (1.33)	-0.1603 (2.47)	0.0900 (1.04)
β_5	0.5405 (3.24)	0.5401 (3.29)	0.5392 (12.71)	0.4295 (3.74)
μ	0.0804 (0.26)	0.0865 (0.43)	0.5129 (2.18)	*c*
Constant	2.7780 (3.28)	2.4480 (3.25)	0.4180 (1.78)	0.4285 (0.74)
SSR	168.31	168.31		81.08
R^2	.175[d] .212[e]	.175[d] .215[e]	.158[d]	.134[d] .199[e]
F	3.278	3.521	1822.6	3.826

[a]Using equation (1), Table 2, *supra*, p. 21, for the first-stage estimates.

[b]Number in parentheses is the absolute value of the t statistic.

[c]Not applicable.

[d]Square of simple correlation coefficient between actual and predicted number of visits.

[e]Square of simple correlation coefficient between actual and predicted number of visits but using log of number of visits.

$\mu = 0.5$, together with some standard error. Theil and Goldberger suggest a standard error of 0.25, which is used here. The resulting coefficient estimates are shown in the second column of Table 3. They are very similar to the coefficients in the first column; the estimate of μ is slightly higher but still not statistically significant.[20]

The constant term in the first two columns of Table 3 is an estimate of $\ln [\theta_0 \Gamma (1 - \mu)]$. However, Goldberger [8] and Heien [13] have pointed out that, if one takes the exponential of this estimate, e.g., $e^{2.448} = 11.5626$ from the second column, it yields a <u>biased</u> estimate of $[\theta_0 \Gamma(1 - \mu)]$. Accordingly, I employed Goldberger's correction factor, which was computed to be 2.2605.[21] Therefore, I take 11.5626 x 2.2605 = 26.1375 as my estimate of $[\theta_0 \Gamma(1 - \mu)]$. Given the estimate of $\mu = 0.0865$, the implied estimate of θ_0 is 24.7058.[22] Using the same estimate of μ, one can obtain estimates of the parameters γ_k and h from the formulas $\gamma_k = \bar{\gamma}_k \cdot \mu$ and $h = \bar{h} \cdot \mu$, where $\bar{\gamma}_k$ and $\bar{h}$ are the first-stage estimates presented in the first column of Table 2. The resulting final coefficient estimates are shown in the first column of Table 4.[23]

The third column of Table 3 contains estimates of (22) based on nonlinear least squares applied to the full sample. The coefficient estimates are qualitatively similar to those in the first two columns of the table, except that the estimate of β_4 is now negative. The estimate of μ is 0.5129 and is statistically significant. The estimate of the constant term, which here represents $[\theta_0 \Gamma(1 - \mu)]$, is 0.418. Taking $\mu = 0.5129$, this implies an estimate of 0.2299. For θ_0, the implied estimates of γ_k and h are shown in the second column of Table 4.[24]

TABLE 4

Final Estimates of the Coefficients of the Blackburn Model

Coefficient	Estimates[a]	
	Model 1[b]	Model 2[c]
γ_1	- 0.0025 (0.41)	- 0.015 (1.85)
γ_2	- 1.4513 (0.41)	- 8.6095 (1.80)
γ_3	- 0.1845 (0.42)	- 1.0945 (1.96)
γ_4	- 0.0459 (0.36)	- 0.2725 (1.25)
γ_5	0.1353 (0.40)	0.8027 (1.59)
h	0.1184 (0.43)	0.7022 (2.14)
θ_0	24.7058	0.2299

[a]The number in parentheses is the absolute value of the t statistic.

[b]Based on equation (2) of Table 3, *supra*, p. 23; the estimates of the other coefficients of the model are given in that table.

[c]Based on equation (3) of Table 3, *supra*, p. 23; the estimates of the other coefficients of the model are given in that table.

The overall fit of these two models can be assessed by computing the predicted total number of visits to each site by the households in the sample and comparing this with the actual total. The square of the simple correlation coefficient between the actual and predicted totals for the 20 sites is 0.407 for the model in the first column of Table 4 and 0.433 for the model in the second column. The square of the simple correlation coefficient between the actual and predicted number of visits by each household to the particular site which it actually selected is 0.175 for the first model and 0.365 for the second model. By these criteria, the second model--estimated from the full sample--is to be preferred; it will be used for the consumer's surplus calculations presented in the next section.

Suppose one did not know of the qualitative-quantitative choice models such as (7) and (8). How else might one have analyzed the beach visitation data I shall briefly describe an alternative ad hoc demand model, not derived from any utility maximization hypothesis, which one might think of employing in these circumstances. I will compare its predictive power with that of the two models estimated above. As before, let $\tilde{x}_i$ be the number of visits by a household to site i. The model is

$$E\{\tilde{x}_i\} = \Pr\{\text{household selects site } i\} \cdot E\{\text{total number of recreation trips by household}\}. \tag{25}$$

The model is superficially similar to (15) and (22), but there are important differences. The first term on the right-hand side of (25) is equivalent to P^i, as given by (13), and can be estimated by applying MNL to the qualitative choice data. The second term on the right-hand side of (25) comes from a

regression of the total number of recreation trips by the household on various explanatory variables and can be formulated exactly as in the right-hand side of (24). Thus, (25) differs from (22) by the omission of the term $\bar{\lambda}^{\mu}$.[25] For the qualitative choice probabilities, I use the coefficient estimates in the first column of Table 2; i.e., I use the MNL model in standard form, with $\mu = 1$, based on the generic specification (23). For the second part of the model, I regress the total number of recreation trips by each household on the right-hand side of (24), taking the logarithm of both sides and applying OLS. The results are shown in the last column of Table 3.[26] Then, using (25), I predict the number of trips by each household to each site and sum this over all households. The square of the simple correlation coefficient between the actual and predicted total number of visits to each site is 0.413. The square of the simple correlation coefficient between the actual and predicted number of visits by each household to the particular site which it actually selected is 0.088. By these criteria, one may conclude that this ad hoc demand model is inferior in its predictive power to the qualitative-quantitative choice model in the second column of Table 4.

5. BENEFIT ESTIMATION

In this section the fitted qualitative-quantitative choice model in the second column of Table 4 will be used to calculate two sets of benefit measures. The first set deals with the benefits from improving water quality. I simulated the effects of reducing either COD or PHOS at each site taken separately; I considered both a 10 percent and a 50 percent reduction in these variables. The benefit to each household was calculated from formula (17). Since households vary in their location and socioeconomic characteristics,

they do not receive the same benefit from a given reduction in COD or PHOS at a particular site. To save space, only the average benefits per household are presented in Table 5. Thus, the first entry states that the average benefit per household from a 10 percent reduction in COD at site 1 is 12.3 cents. In fact, the minimum benefit to a household is 0.7 cents, and the maximum is $1.57--the standard deviation is 21 cents. A similar variation underlies the other entries in the table. Note that these numbers are not the benefit per visit but the total benefit over the whole summer recreation season--the number of visits by each household to the recreation site is already incorporated in formula (17). Note, also, that the benefits from a 50 percent reduction in pollutants are generally more than five times the benefits from a 10 percent reduction; this reflects the nonlinearity of the underlying demand model.

In addition to generating estimates of the benefits from a change in site quality, the qualitative-quantiative choice model can be used to generate an estimate of the consumer's surplus from each site, i.e., the benefit from the mere existence of the site analogous to the Marshallian triangle under a demand curve. The idea behind this calculation is that, if a site were to become unavailable to a household, this would be equivalent to an increase in its price from the current level to a level at which the household's demand for the site would fall to zero. One measures the benefit to the household from the existence of the site as the benefit of a price change from the zero-visitation level to the actual level for that household, using (17). There is one qualification: in the Blackburn model a price of infinity would be required to drive a household's expected demand for a site to zero--in effect, the demand curve is asymptotic to the vertical axis. Therefore, I use a cut-off price such that each household's probability of visiting the site

TABLE 5

Average Benefit Per Household from a 10 Percent and 50 Percent Reduction in Water Pollutants Based on the Blackburn Model[a]

Site	Benefit from 10 Percent Reduction in:		Benefit from 50 Percent Reduction in:	
	COD	PHOS	COD	PHOS
	cents			
1	12.3	9.7	72.1	54.9
2	11.9	4.6	83.6	26.1
3	9.1	8.2	54.7	48.1
4	5.4	8.4	32.7	57.2
5	8.6	8.2	53.1	50.4
6	2.6	7.0	16.1	60.2
7	3.5	2.2	29.9	15.0
8	6.7	6.2	49.4	44.3
9	19.5	13.2	128.3	79.6
10	6.0	6.4	38.8	42.5
11	1.2	3.8	7.7	36.4
12	*b*	13.7	*b*	77.1
13	2.0	1.8	11.7	10.2
14	3.8	1.4	22.6	7.3
15	2.4	0.5	17.2	2.5
16	1.5	1.2	9.8	7.1
17	0.9	0.4	5.6	2.5
18	1.0	2.0	5.7	13.8
19	0.8	1.2	4.2	6.3
20	*b*	1.6	*b*	9.3

[a]Based on Model 2, Table 4, *supra*, p. 25.

[b]No benefit since COD already is zero at the site.

falls to an arbitrarily small (but nonzero) level.[27] The results are shown in the first column of Table 6. These figures are the average consumer's surplus per household from each site. The first entry in the table states that the average consumer's surplus from site 1 is $1.45 per household. The minimum for this site for any household is 7.6 cents, and the maximum is $18.94; the standard deviation is $2.46. The mean consumer's surplus per household over all the sites is about 54 cents. Note that these figures are consumer's surplus per summer recreation season. Most previous estimates of the consumer's surplus from recreation sites have been couched in terms of the benefit per household visit or per visitor day. In order to compare my results with these estimates, I divided the total consumer's surplus for each site by the predicted total number of visitor days at the site, using the total number of household visits at each site predicted by the model and survey data on the average number of persons in each household's party when it visits a site. The results are shown in the second column of Table 6. The average over all sites is 17.4 cents per visitor day.[28]

Finally, suppose that one had employed the ad hoc demand model, (25), instead of (12). How might one calculate benefit measures from price or quality changes with this model The logic of the model is that price or quality changes influence site selection probabilities--the first term of the right-hand side of (25)--but not total household recreation activity--the second term in (25). Consider, first, the effects of a change in the quality of a site. Each time a household selects a recreation site, i.e., each time it makes a recreation trip, it reaps some benefit from the change; this benefit is given by formula (21). Over the entire summer recreation season, the total

TABLE 6

Average Consumer's Surplus Per Household and Per Visitor Day
For Sites in the Boston Area

	Blackburn Model[a]		Ad Hoc Model[b]	
	Consumer's Surplus Per:		Consumer's Surplus Per:	
Site	Household	Visitor Day	Household	Visitor Day
	dollars	cents	dollars	cents
1	1.45	17.7	0.89	18.8
2	0.66	17.2	0.41	17.8
3	0.95	17.0	0.59	17.6
4	0.51	16.6	0.34	16.7
5	0.79	17.5	0.57	17.1
6	0.24	18.5	0.15	17.3
7	0.13	16.4	0.09	15.4
8	0.34	17.2	0.23	16.1
9	1.32	20.4	0.87	19.4
10	0.45	17.9	0.33	16.4
11	0.11	17.1	0.09	15.7
12	2.17	21.5	2.22	25.1
13	0.25	17.6	0.14	18.2
14	0.40	17.6	0.21	18.6
15	0.13	16.6	0.07	17.1
16	0.10	19.9	0.11	20.9
17	0.05	18.5	0.06	19.2
18	0.12	14.1	0.06	15.3
19	0.34	12.9	0.16	14.8
20	0.18	15.6	0.09	14.2

[a]Based on Model 2, Table 4, *supra*, p. 25.

[b]For explanation, see *infra*, p. 32.

benefit to the household is obtained by multiplying this benefit per site choice by the household's total number of recreation trips for the season. A similar procedure may be employed for calculating the consumer's surplus from a site defined, as above, as the household's benefit from a hypothetical price change from the zero-visitation level to the actual level for that household. Again, I use a cutoff price such that the household's probability of visiting the site falls to an arbitrarily small but nonzero level. Application of formula (21) yields the consumer's surplus per site choice, and multiplication by the household's total number of recreation trips yields the total consumer's surplus over the summer recreation season. This exercise was performed using the coefficient estimates in the first column of Table 2 and the actual total number of recreation trips by each of the 83 households. The average household's consumer surplus for each site is tabulated in the third column of Table 6. The first entry in the column shows that the average consumer's surplus from site 1, using the ad hoc demand model, is 89 cents per household. The average over all the sites is about 38 cents per household. These figures may be converted to a per visitor day basis by dividing the total consumer's surplus for each site by the actual total number of visitor days at the site. The results are shown in the last column of Table 6; the average consumer's surplus over all sites is 17.6 cents per visitor day.

Thus, although the ad hoc demand model is conceptually distinct from the Blackburn model, it yields similar estimates of consumer's surplus per visitor day.[29] It is impossible to make a precise comparison of these estimates with the other estimates which have appeared in the recreation demand literature because of differences in the time and geographical location of the

study, the type of water-based recreation activity encompassed, and the methodology used for modeling demand and estimating consumer's surplus. However, a rough comparison may be made. The project evaluation guidelines promulgated by the Water Resources Council in 1973 specify a range of values of 75 cents to $2.25 per visitor day for general water-based recreation [23]. This range is broadly consistent with the findings of recreation demand studies reported in the literature. In [10], 11 studies are summarized, conducted mainly in the 1960s and early 1970s, which generally yielded consumer's surplus estimates either in this range or higher. In particular, Burt and Brewer's [3] estimate of consumer's surplus for three hypothetical reservoirs near St. Louis comes to $2.43 per visitor day. If my results for the Boston area can be extrapolated elsewhere, they would suggest that the values per visitor day now widely used for recreation benefit evaluations are significantly exaggerated.

6. CONCLUSIONS

The purpose of this paper has been to describe some empirical demand models which can be applied when a consumer makes both a qualitative choice—which one of N items to select—and a quantitative choice—how much of the chosen item to buy. Standard quantitative demand models, such as the linear expenditure system, cannot do justice to the qualitative nature of these choices. Quantal choice models, such as MNL, cannot accommodate their quantitative aspect. The Tobit model and its generalizations capture both features but do not integrate them into a common utility maximization framework. Therefore, the demand models described here fill a gap in the literature. Moreover, by deriving these models from an explicit utility maximization problem, from the fitted demand functions I can construct exact welfare measures

for changes in the prices or qualities of the items in the choice set. The Blackburn model on which I focus in the empirical sections of the paper is restrictive in that it implies zero income elasticities of demand for the branded items. In this particular application, however, this may not be an unreasonable restriction: the results in [4], as well as my own data, provide strong evidence that the demands for individual recreation sites are not responsive to household income.[30] The other demand model described in this paper avoids this restriction, but it is more complicated to estimate. The main substantive results of the empirical application are the finding that certain aspects of water and beach quality at recreation sites significantly influence recreation choices, and the estimates of consumer's surplus which, if they can be extrapolated to other urban areas, imply strongly that the Water Resources Council's guidelines for valuing a general recreation day are excessive.

University of California, Berkeley

APPENDIX

The sites included in the empirical analysis (and their location) are:

1. Lynn Beach (Lynn).
2. Nahant Beach (Nahant).
3. Revere Beach (Revere).
4. Constitution Beach/Orient Heights (Boston).
5. Castle Island (Boston).
6. City Point (Boston).
7. L&M Street Beaches (Boston).
8. Carson Beach (Boston).
9. Malibu Beach/Savin Hill (Boston).
10. Tenean Beach (Boston).
11. Wollaston Beach (Quincy).
12. Nantucket Beach (Hull).
13. Wingaersheek Beach (Gloucester).
14. Crane's Island (Ipswich).
15. Plum Island (Newberry).
16. Duxbury Beach (Duxbury).
17. White Horse Beach (Plymouth).
18. Wright's Pond (Medford).
19. Walden Pond (Concord).
20. Cochituate State Park (Natick).

The last three are freshwater beaches; the others are all saltwater ocean beaches.

FOOTNOTES

[1] I wish to thank Professors Dale Jorgenson and Robert Dorfman for their generous assistance in supervising the dissertation on which this paper is based. I am grateful to the U. S. Environmental Protection Agency for funding the collection of the data used in the empirical portion of the paper.

[2] For an excellent survey of quantal choice models, see [18].

[3] On the latter point, see [16].

[4] A tilde will be used to denote a random variable or function.

[5] Another utility function with the same property is

$$u\ (x,\ \tilde{\phi},\ z) = z^{\alpha}\ (\Sigma\, x_j)^{\beta} - \Sigma\, \tilde{\phi}_j\ \delta(x_j), \qquad \delta(s) \begin{cases} = 1 & \text{if} \quad s > 0 \\ = 0 & \text{if} \quad s = 0 \end{cases}$$

which appears as a numerical example in [20].

[6] The domain of $\psi(\cdot)$ depends on the specification of $g(\cdot)$; its range is $(0,\ \infty)$.

[7] It should be noted that the decomposition does <u>not</u> apply to the general "ruled indifference surface" model presented in [20] in which $\tilde{\phi}_j = \tilde{\phi}_j\ (b_j,\ z)$.

[8] That is, $\Pr\{\tilde{\varepsilon}_j \leq s\} = \exp\ \{-\exp\ [-(s - \alpha_j)/\mu]\}$.

[9] In [1], the function $g(\cdot)$ itself contains a stochastic term and is defined by $\tilde{g}(s) = s\ [1 + \ln\ \theta - \tilde{\beta} - \ln\ s]$. In [2], the random coefficient $\tilde{\beta}$ is eliminated. I shall focus on the latter model.

[10]Blackburn never explicitly presents this formula for his utility function; instead he gives a formula for $\tilde{\pi}_j$ similar to (12b) and a formula for $\psi(\cdot)$ from which $g(\cdot)$ can be obtained by integration. There are some small differences between Blackburn's implicit utility function and (12): his $\gamma_j(b_j)$ is the negative of mine, and he makes h and θ functions of characteristics of the individual consumer. The latter modification is introduced in the next section.

[11]Compare (13) and (15) with equations (3.17) and (3.14) of [2], noting that (i) Blackburn's C_j corresponds to my π_j; (ii) his λ corresponds to my h; (iii) he writes $\mu = h/\xi$, where ξ is the coefficient to be estimated; and (iv) his $\gamma_j(b_j)$ is the negative of mine. Note that he actually writes the gamma term in (15) as $\Gamma[(h/\xi) - 1]$ instead of $\Gamma[1 - (h/\xi)]$, which appears to be an error.

[12]The question of welfare measures for price or quantity changes is not addressed in [1, 2]. Note that, in the present case, the range of $\tilde{u}_i$ is (hy, ∞).

[13]An additional normalization is required in estimating (13) since it is invariant to multiplication of both the numerator and denominator by e^{δ}, for an arbitrary constant δ. An appropriate normalization would be to set $\alpha_j = 0$ for one index j.

[14]This can be shown to be equivalent to equation (4.103) in [5].

[15]This procedure yields biased estimates of the constant term; this bias and the method of correcting it are discussed in the next section.

[16]The 106 households visited 30 sites altogether. Eight relatively less important sites had to be omitted because I had no data on their water quality, which left 22 sites. The TROLL program for estimating the MNL model which I used could only handle $N \leq 20$. Therefore, two more sites were omitted.

[17]Other water and nonwater quality attributes were tested but were found to be highly collinear with these variables. Details of these tests and of other estimation results not described here are provided in [10, chapter 7].

[18]Unlike some recreation demand studies, I did not include a time component in travel cost. The data do not indicate how many adults or children were in the party visiting the site, nor whether the trips were made on weekdays or the weekend. It seems impossible to derive any reasonable estimate of the shadow price of the time spent visiting the sites.

[19]The hypothesis that the coefficients β_4 and β_5 are equal could be rejected at the .9 level but not at the .95 level.

[20]The hypothesis that the prior and sample information are mutually consistent was tested using Theil's [21] compatability statistic and could not be rejected at the 0.9 level.

[21]In terms of Goldberger's notation, here $s^2 = 2.2146$, $m^{00} = 0.2558$, $cw = 0.82408$, and $v = 76$.

[22]Since this estimate is a highly nonlinear function of the estimates of μ and the other coefficients, I have not attempted to approximate an estimate of its variance.

[23]The variances of these estimates and, hence, the t statistics are calculated from Goodman's [9] formula: $V(xy) = E^2(x)\ V(y) + E^2(y)\ V(x) + V(x)\ V(y)$. The formula is applicable only if the estimates of $\bar{\gamma}_k$ or $\bar{h}$ and μ are stochastically independent; this is not strictly true since the estimates of $\bar{\gamma}_k$ and $\bar{h}$ are inputs to the estimation of μ. Nevertheless, it seems reasonable as an approximation. The large variance of the estimate of μ is responsible for the relatively low t statistics appearing in the table.

[24]The t statistics are calculated in the same way as for the first column of the table.

[25]Apart from the fact that there are 20 alternatives (sites) instead of 2, this model is a special case of the general limited dependent variable model presented in [17, p. 358]. In the notation of that paper, I am assuming that $\beta_1 = \beta_2$, which results from the generic specification of site demand, and $\sigma_{1\epsilon} = \sigma_{2\epsilon} = 0$.

[26]For the same reasons as before, the estimate of the constant term in the table must be adjusted by Goldberger's correction factor. In this case, $s^2 = 1.0668$, $m^{00} = 0.31212$, $cw = 0.36692$, and $v = 76$; the correction factor is computed to be 1.4408.

[27]I use a cutoff probability of 0.0008, which is the smallest value in the matrix of predicted site visitation probabilities. The results are not very sensitive to this choice of a cutoff probability; some experiments show that reducing it by a factor of 10 changes the average consumer's surplus for a site by well under 1 cent.

[28]For some sites, the estimated consumer's surplus per household is less than the consumer's surplus per visitor day. This occurs because the model predicts a total season demand of less than one visitor day for these sites.

[29]The simple correlation coefficient between the first and third columns of Table 6 is 0.961; that between the second and fourth columns is 0.843.

[30]This is after taking family size into account, which is correlated with household income.

REFERENCES

[1] BLACKBURN, A. J.: "A Non-Linear Model of the Demand for Travel," Chapter 8 in The Demand for Travel: Theory and Measurement, ed. by R. E. Quandt. Lexington: D. C. Heath, 1970.

[2] _____: "An Alternative Approach to Aggregation and Estimation in the Non-Linear Model," Chapter 9 in The Demand for Travel: Theory and Measurement, ed. by R. E. Quandt. Lexington: D. C. Heath, 1970.

[3] BURT, OSCAR R., and DURWARD BREWER: "Estimation of Net Social Benefits from Outdoor Recreation," Econometrica, 39 (1971), 813-827.

[4] CICCHETTI, CHARLES, ANTHONY C. FISHER, and KERRY V. SMITH: "An Econometric Evaluation of a Generalized Consumer Surplus Measure: The Mineral King Controversy," Econometrica, 44 (1976), 1259-1276.

[5] DOMENCICH, THOMAS A., and DANIEL McFADDEN: Urban Travel Demand: A Behavioral Analysis. Amsterdam: North-Holland, 1975.

[6] DUNCAN, GREGORY M: "Formulation and Statistical Analysis of the Mixed Continuous/Discrete Dependent Variable Model in Classical Production Theory," Econometrica, Vol. 48, No. 4 (May, 1980), 839-851.

[7] FROME, EDWARD L., MICHAEL H. DUTNER, and JOHN J. BEAUCHAMP: "Regression Analysis of Poisson-Distributed Data," Journal of the American Statistical Association, 68 (1973), 935-940.

[8] GOLDBERGER, ARTHUR S.: "The Interpretation and Estimation of Cobb-Douglas Functions," Econometrica, 35 (1968), 464-472.

[9] GOODMAN, L. A.: "On the Exact Variance of Products," Journal of the American Statistical Association, 55 (1960), 708-713.

[10] HANEMANN, W. M.: "A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement," Ph.D. dissertation, Department of Economics, Harvard University, 1978.

[11] HAUSMAN, JERRY A., and DAVID A. WISE: "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," Econometrica, 46 (1978), 403-426.

[12] HECKMAN, J.: "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, 5 (1976), 475-492.

[13] HEIEN, DALE M.: "A Note on Log-Linear Regression," Journal of the American Statistical Association, 63 (1968), 1034-1038.

[14] JORGENSON, D. W.: "Multiple Regression Analysis of a Poisson Process," Journal of the American Statistical Association, 56 (1961), 235-245.

[15] LANCASTER, KELVIN: Variety, Equity, and Efficiency. New York: Columbia University Press, 1979.

[16] LAU, L. J.: "Complete Systems of Consumer Demand Functions Through Duality," in Frontiers of Quantitative Economics, Vol. III, ed. by M. D. Intriligator. Amsterdam: North-Holland, 1977.

[17] LEE, LUNG-FEI, and ROBERT P. TROST, "Estimation of Some Limited Dependent Variable Models with Application to Housing Demand," Journal of Econometrics, 8 (1978), 357-382.

[18] McFADDEN, DANIEL: "Quantal Choice Analysis: A Survey." Annals of Economic and Social Measurement, 5 (1976), 363-390.

[19] _____: "Modelling the Choice of Residential Location," in Spatial Interaction Theory and Planning Models, ed. by Anders Karlquist, Lars Lundqvist, Folke Snickars, and Jorgen W. Weibull. Amsterdam and New York: North-Holland, 1978.

[20] NOVSHEK, WILLIAM, and HUGO SONNENSCHEIN: "Marginal Consumers of Neoclassical Demand Theory," Journal of Political Economy, 87 (1979), 1368-1376.

[21] THEIL, H.: "On the Use of Incomplete Prior Information in Regression Analysis," Journal of the American Statistical Association, 58 (1963), 401-414.

[22] THEIL, H., and A. S. GOLDBERGER: "On Pure and Mixed Statistical Estimation in Economics," International Economic Review, 2 (1961), 65-78.

[23] Water Resources Council: "Water and Related Land Resources," Federal Register, 38 (1973).