*Journal of Agricultural and Resource Economics* 21(2):210–219


# Hurdle Count-Data Models in Recreation Demand Analysis

**J. S. Shonkwiler and W. Douglass Shaw**

When a sample of recreators is drawn from the general population using a survey, many in the sample will not recreate at a recreation site of interest. This study focuses on nonparticipation in recreation demand modeling and the use of modified count-data models. We clarify the meaning of the single-hurdle Poisson (SHP) model and derive the double-hurdle Poisson (DHP) model. The latter is contrasted with the SHP and we show how the DHP is consistent with Johnson and Kotz's zero-modified Poisson model.

*Key words*: count-data models, discrete probability, hurdle count-data models, recreation demand

## Introduction

Recreation demand modeling has been used by economists since the early 1960s, when economists began to use travel costs to reveal a recreator's preferences (Smith). Increasingly, sophisticated empirical models are developed for individual recreators rather than for aggregate groups of individuals. Two of the most popular individual models are the random utility model (RUM) and the count-data model.[1]

Data on trips individuals take to recreation sites are usually recorded using a survey instrument. When the data include the total number of trips taken in a given period (year or recreation season), and especially when these data are only available for one recreation site, the count-data model is attractive. Alternatively, if the survey questionnaire obtains detailed information on each trip (e.g., where, when, length of stay, activities) to many recreation destinations, the RUM model can be quite useful.

This study focuses on the count-data model's treatment of responses recorded as zero. We are especially interested here in nonparticipation (i.e., when many of those surveyed take no trips at all to one site or to any site in a group of recreation sites). Only recently have economists been able to make inferences about those who make no trips. As shown by Morey, Shaw, and Rowe, allowing individuals to substitute out of recreation altogether affects welfare estimates. For example, a particular environmental consequence may be so severe that those, who would normally recreate at the site, discontinue their activities. Hellerstein (1992) reviews how nonparticipants have typically been treated in travel cost analyses. He describes how data sets with individuals providing censored trip data (non-

Authors are, respectively, professor and associate professor in the Department of Applied Economics and Statistics at University of Nevada, Reno.

Research supported by the Nevada Agricultural Experiment Station. The authors thank Richard Carson for his comments on this manuscript. We also thank three anonymous referees, whose comments have greatly helped us improve this paper.

[1] Bockstael, McConnell, and Strand review the modern approaches in recreation demand modeling. Recent recreation applications of the RUM with a nested logit model are in Parsons and Needelman and Morey, Shaw, and Rowe. Applications with the count-data model are Englin and Shonkwiler, Hellerstein (1991), and Creel and Loomis.

negative, but including zero trips) are often converted into truncated data sets by dropping all those who take zero trips. He notes that the Poisson count-data model allows for zero outcomes and censoring, but that modifications must be made for truncation. Issues involving other types of corner solutions are discussed in Morey et al.

Count-data estimators have now been considered which allow for the fact that the data are truncated due to sampling on-site (Creel and Loomis; Shaw; Englin and Shonkwiler). Hellerstein (1992) briefly refers to two-stage count-data models, where the transition from zero to positive trips is handled separately from the probability of choosing the exact number of positive trips. What he describes is essentially Mullahy's single-hurdle model.

Though the microeconomic foundations for deriving welfare measures under conditions of discrete demand have been addressed (Hellerstein and Mendelsohn), there appears to be confusion regarding interpretation of the welfare measures using modified forms of the count-data models because the connection between the underlying probability mass function and economic behavior is often not established (Yen and Adamowicz). We define a "user" as a person who is currently recreating and a "nonuser" as a person who has not recreated in the past, is not now, and likely will not recreate in the future. We define a "potential user" as a person who is not currently recreating but who might begin recreating with a change in the price of recreating. In steak consumption, the analogous groups of individuals would be the steak eater (user), the vegetarian (nonuser), and the hamburger eater (potential user). The important distinction to keep in mind is that the vegetarian will not consume steak even under a drastic price decrease, while the hamburger eater may start consuming steak under a price decrease or income increase.

We concentrate here on the so-called hurdle count-data models (Mullahy). We demonstrate that by using the double-hurdle (DH) model we can explore the behaviors of all three types of individuals defined above. While several studies use a DH model to model behavior when nonzero responses are continuously distributed, apparently no studies have examined responses generated under a discrete distribution. Further, while some have discussed "zero" modifications to the Poisson (e.g., Greene), no one has made the connection between these and the DH model. We first carefully lay out the probability mass function (PMF) for the simple Poisson model. Next, we do the same for the single-hurdle model and the double-hurdle model assuming no interdependence between the hurdle mechanisms. We then develop a double-hurdle model allowing for interdependence between the hurdles.

## The Probability Mass Function in the Basic Count-Data Model

Assume that the $i$th potential user of a specific recreation site has been randomly drawn from a relevant population. The relevant population is the population of users plus potential recreators. Let $y_i$ denote the number of visits to the recreation site made by the individual. If the probability function for $y_i$ is

$$\text{(1)} \qquad \text{Prob}(Y_i = y_i) = e^{-\lambda_i} \lambda_i^{y_i}/y_i!, \qquad y_i = 0,1,2,3, \ldots,$$

where $Y_i$ is a potential integer outcome, then it is well known that $E(Y_i) = \lambda_i$.

Assume we randomly sample the relevant population to obtain information on $y_i$. As Hellerstein (1992) notes, different survey methods can lead to different types of individuals, and one common approach in recreation surveys is to screen the general population to obtain a sample of the recreators. If the final random sample consists only of users (known recreators), then $q_i = y_i | y_i > 0$. The use of $q_i$ denotes that the underlying population of all users, including *potential* users from which $y_i$ was drawn, has been truncated. The sample $q$, of which $q_i$ is an element, only provides information about the population of users; it tells us nothing about the population of users *and* potential recreators unless the probability mass function in (1) has been empirically verified. Since $q$ is a subset of $y$ (the sample of the population of all known and potential recreators), knowledge of $y$ can always be used to make inferences concerning known users; however, as is the typical case in sample inferences, knowledge of $q$ cannot be used to make inferences about known and potential recreators because there is no means to establish (1) as the underlying probability mass function.

The probability mass function for the $j$th individual in the sample of observed recreators given the Poisson parameter $\psi$ is

$$(2) \qquad \text{Prob}(Q_j = q_j) = \frac{e^{-\psi_j}\dfrac{\psi_j^{q_j}}{q_j!}}{1 - e^{-\psi_j}} \qquad q_j = 1, 2, 3, \ldots,$$

where again $Q_j$ is defined so that it is an integer greater than zero and $E(Q_j) = \psi_j/(1 - e^{-\psi_j})$. Expression (2) is called the positive Poisson distribution. Because the denominator in (2) can be interpreted as the probability that $q_j$ is greater than zero, some have interpreted $e^{-\psi_j}$ as the probability of nonuse. The discussion of Grogger and Carson (p. 230) regarding this issue fails to stress the assumption that it is known with certainty that the underlying distribution follows the Poisson with the identical location parameter ($\lambda_j = \psi_j, \forall_j$). In general, the sample $q$ can tell us nothing about potential recreators because it contains no information about them. This probability, $e^{-\psi_j}$, has no economic interpretation relating to decisions to start recreating or stop. Instead, the denominator in (2) accounts for the correction necessary when $Q_j$ is prohibited from being zero so that (2) is a valid probability mass function.

## Hurdle Count-Data Models

The basic Poisson travel cost model accommodates zero trips. However, if the sample is drawn from the general population which includes those who are recreators but for some reason did not recreate in the past season (potential users), as well as those who *never* have and never will recreate (nonusers), then it is quite likely that many or most individuals take no trips at all to a particular recreation site.

Handling data sets which include many individuals who do not participate can be difficult. Some have used the negative binomial distribution rather than the Poisson in an attempt to handle the "excess zeros." Modifications include Lambert's zero-inflated Poisson (ZIP) model, implemented to adjust for the situation where the Poisson underpredicts the zero values (Greene). However, the hurdle models (Mullahy) do this in a manner which is perhaps more intuitive and allow decomposing the decision processes underlying ($a$) the choice to recreate at all and ($b$) the choice to recreate $n$ times at a certain site. The hurdle model has also been deemed the "zero-altered Poisson" or ZAP model (see Greene's discussion). Pudney suggests that the factors explaining the two decisions can be separated into economic variables (travel cost and site quality) and personal characteristics (physical or athletic ability, household size, marital status, gender, etc.).

To develop a single-hurdle model, assume that $y$ represents a general population sample containing known users, potential users, and nonusers. Let $y_i$ denote the number of visits individual $i$ has made to a site over the course of some period (a season). Define two vectors of variables, $x$ and $z$, where $x$ contains economic variables most likely bearing on the decision to visit a given site $n$ times, and $z$ contains personal or demographic variables pertaining to the decision to recreate at all during the season.[2] Let $D_i$ represent the **latent** decision to recreate at a site. The quantity of recreation at that site is zero if the random variable $D_i \leq 0$.[3] As is conventional with a latent variable approach, $D_i$ can be negative. Hellerstein (1992) notes that negative recreational visits (demand) are impossible, but we stress here that $D_i$ merely indicates whether there are unobserved impediments which preclude the individual from visiting the site during the season.[4] Specifically, we adopt the discrete specification $\text{Prob}(D_i = 0) = \exp(-\theta_i)$ where $\theta_i$ can be parameterized

$$\theta_i = \exp(z_i'\gamma), \tag{3}$$

and $\gamma$ is an unknown vector of parameters.[5]

If consumption is positive, then observed consumption equals desired consumption, $y_i^*$, so that

$$y_i = y_i^* \tag{4}$$

with

$$E(y_i^*) = \lambda_i = \exp(\alpha_i + \beta p_i),$$

where $p$ is the travel price, $\beta$ is its corresponding parameter, and $\alpha_i$ represents socioeconomic measures associated with the $i$th individual. The sequential (single-hurdle) model then has a dichotomous probability mass function of the form:

$$\begin{aligned} &\text{Prob}(D_i \leq 0); && \text{if } y_i = 0, \\ \text{PMF}(y_i | y_i > 0)&\text{Prob}(D_i > 0); && \text{if } y_i > 0, \end{aligned} \tag{5}$$

which implies that $\text{Prob}(y_i > 0) = 1 - \exp(-\theta_i)$. The likelihood function in the case of the single-hurdle model with Poisson PMF specification is (Mullahy)

$$\frac{\prod_{y=0} \exp(-\theta_i) \prod_{y>0}(1 - \exp(-\theta_i))\exp(-\lambda_i)\lambda_i^{y_i}}{[(1 - \exp(-\lambda_i))y_i!]}, \tag{6}$$

or

[2] Note that it is certainly possible that some variables appear in both vectors. Examples of variables that might influence the two decisions are mentioned by Pudney (see above) and are offered in the conclusion of this article.

[3] The distribution of $D_i$ may be continuous and defined in the interval $(-\infty,\infty)$ or discrete and defined over $[0, 1, 2, \ldots]$. In the latter case, consumption is zero if $D_i = 0$.

[4] We thank an anonymous referee for forcing us to clarify this point.

[5] It is quite possible another approach may better reflect the data generating mechanism. For example, if $D_i$ is taken as normally distributed, then $\text{Prob}(D_i \leq 0) = \Phi(-z_i'\gamma)$, where $\Phi$ is the cumulative distribution function for the normal distribution.

$$\frac{\prod_{y=0} \exp(-\theta_i) \prod_{y>0} (1 - \exp(-\theta_i))\lambda_i^y}{[(\exp(\lambda_i) - 1)y_i!]}.$$

This formulation is attractive because if $\theta_i = \lambda_i$, then the single-hurdle collapses to a simple Poisson specification. For this model,

$$(7) \qquad E(y_i | y_i > 0) = \frac{\lambda_i}{1 - e^{-\lambda_i}} \quad \text{and} \quad E(y_i) = \frac{\lambda_i(1 - e^{-\theta_i})}{1 - e^{-\lambda_i}}.$$

This model implies that the decision to consume an additional trip to a site is independent of the probability of observing any trips to that site.[6] This is a rather strong assumption, as it implies that a recreator's preferences for another (marginal) trip have nothing to do with the probability of observing a positive number of trips to a recreation destination.[7] In addition, while this single-hurdle model is empirically tractable and leads to an expression for expected consumer's surplus, there are certain other disadvantages to using it to model recreation demand which will be overcome with the double-hurdle model. To foreshadow the main disadvantage, note that (6) above generates probabilities of zero trips with a single mechanism. It cannot tell us why this is so, and therefore, the possibilities that the individual is a nonuser or that the individual is a potential user cannot be inferred.

## Development of the Double-Hurdle Model with Independent Hurdles

Another hurdle model can be developed which allows for two ways of generating zero trips. This model, the double hurdle, splits the decision into one part which is fundamentally noneconomic, and one which is the usual "no-trip" corner solution in recreation demand (Pudney; Morey et al.). There may be some confusion raised by our suggestion that use of the double hurdle in a count-data model is novel. For example, the title of their article and the fact that they use a negative binomial distribution suggest that Lin and Milon applied the DH approach. However, as they state (p. 726) "empirical analysis employed a count-data *double-hurdle* model developed by Mullahy." (Our italics.) As noted above, Mullahy's model is actually what we categorize as a single- rather than double-hurdle model.

To begin, we assume there is no interdependence between the two hurdles. The notation from the above continues through in this section. Following in the vein of Blundell and Meghir, the double-hurdle model (without dependence) specifies the probability of a zero observation as

$$(8) \qquad \text{Prob}(y_i^* \leq 0) + \text{Prob}(y_i^* > 0)\text{Prob}(D_i \leq 0).$$

No consumption will be observed if desired consumption is nonpositive, or if desired consumption is positive, an additional hurdle ($D_i \leq 0$) may prevent consumption. The PMF of a positive observation reflects that both desired consumption is positive and the additional hurdle is not limiting consumption so that it is of the form:

[6] This condition ($\partial\text{Prob}(y_i > 0)/\partial\lambda_i = 0$) arises because the information matrix is block diagonal, and thus, equivalent estimates may be obtained using the maximum-likelihood method by separately estimating a binary choice and a truncated Poisson model.

[7] We note, however, that many recreation demand models explicitly or implicitly make this assumption about consumption decisions. Examples are the simple logit, repeated logit, and repeated nested logit models (Morey, Watson, and Rowe).

(9) $$\text{Prob}(y_i^* > 0)\text{PMF}(y_i^*|y_i^* > 0)\text{Prob}(D_i^* > 0).$$

The Poisson likelihood in this case becomes

(10) $$\prod_{y=0} [\exp(-\lambda_i) + (1 - \exp(-\lambda_i))\exp(-\theta_i)] \prod_{y>0} (1 - \exp(-\theta_i))\exp(-\lambda_i)\lambda_i^{y_i}[y_i!]^{-1},$$

under the assumption that $y_i = y_i^*$ if $y_i^* > 0$ and $D_i > 0$. In the case of the double-hurdle model, we have

(11) $$E(y_i|y_i > 0) = \frac{\lambda_i}{1 - e^{-\lambda_i}} \quad \text{and} \quad E(y_i) = \lambda_i(1 - e^{-\theta_i}).$$[8]

It is now easier to see the main advantage of the DH model over the single-hurdle model, as well as over the simple Poisson model.[9] The DH model can provide estimates of three different probabilities of participation in the market which correspond to our three regimes, or types of individuals defined in the introduction. The model can predict, for example, the probability of nonparticipation, $e^{-\theta_i}$ (the chance the $i$th individual is a nonuser), the probability of a corner solution, $(1 - e^{-\theta_i})e^{-\lambda_i}$ (the $i$th individual is a potential user who optimizes by not visiting a particular site), and the probability of a user visiting a site one or more times, $(1 - e^{-\theta_i})(1 - e^{-\lambda_i})$.

In addition, the model has a nice statistical interpretation. In essence, the DH model is the same as the zero-modified Poisson model of Johnson and Kotz. Recognizing that equation (8) can be rewritten as:

$$\text{Prob}(D_i \leq 0) + (1 - \text{Prob}(D_i \leq 0))\text{Prob}(y_i^* \leq 0),$$

and that our equation (9) can be rewritten as:

$$(1 - \text{Prob}(D_i \leq 0))\text{Prob}(y_i^* > 0)\text{PMF}(y_i^*|y_i^* > 0),$$

therefore, the transformed distribution in terms of observables is

$$\text{Prob}(D_i \leq 0) + (1 - \text{Prob}(D_i \leq 0))\text{PMF}(y_i = 0),$$

and

$$(1 - \text{Prob}(D_i \leq 0))\text{PMF}(y_i = j) \qquad j = 1,2\ldots,$$

where PMF is the probability mass function of the original distribution. In fact, by pointing out that all uncentered moments of the modified distribution differ from the corresponding moments of the original distribution by the factor $[1 - \text{Prob}(D_i \leq 0)]$, Johnson and Kotz provide the means for our model to be extended to virtually any distribution of interest, including the binomial, negative binomial, and hypergeometric distributions.

As opposed to the single-hurdle model which only collapses to the original PMF when the process generating zero probabilities coincides with the implied probability of a zero outcome from the truncated PMF, the double-hurdle model reduces to the original PMF when $\text{Prob}(D \leq 0)$ is zero. If we specify $\omega = \text{Prob}(D \leq 0)$ and suppressing the observational index $i$ in this and the sequel, then for a Poisson model of $y$, we have the likelihood being

---

[8] The second relation follows since $\text{Prob}(y_i > 0) = (1 - e^{-\lambda_i})(1 - e^{-\theta_i})$.

[9] In contrast to the single-hurdle model, the probability of observing positive demand is affected by the level of demand, that is, $\partial\text{Prob}(y_i > 0)/\partial\lambda_i = (1 - e^{-\theta_i})e^{-\lambda_i} > 0$.

$$\prod_{y=0} [\omega + (1 - \omega)e^{-\lambda}] \prod_{y>0} (1 - \omega)e^{-\lambda}\lambda^{y}[y!]^{-1},$$

in the case of the double-hurdle model.

To a certain extent it is an empirical issue whether the single- or double-hurdle model provides the better fit of a particular data set. As a consequence, the analyst may consider an unrestricted specification which nests both the single- and double-hurdle models. If $\omega$ represents the probability of excess zeros as implied by the double-hurdle model and $e^{-\theta}$ represents an alternative to $e^{-\lambda}$ for generating probabilities of zero observations as in the single-hurdle model, then the full likelihood for a Poisson count-data model becomes

$$\prod_{y=0} [\omega + (1 - \omega)e^{-\theta}] \prod_{y>0} (1 - \omega)(1 - e^{-\theta})e^{-\lambda}\lambda^{y}[(1 - e^{-\lambda})y!]^{-1}. \tag{12}$$

The restriction of $\theta = \lambda$ yields a double-hurdle model, the restriction $\omega = 0$ yields a single-hurdle model, and restricting $\omega = 0$ and $\theta = \lambda$ yields a simple Poisson count-data model. Also, care must be exercised in specifying $\omega$ and $\theta$ because the likelihood function (12) is not identified when the exogenous variables conditioning $\omega$ and $\theta$ either coincide or one set is the subset of the other.

Rejection of all of the above restrictions suggests a hybrid model in which

$$E(y|y > 0) = \frac{\lambda}{1 - e^{-\lambda}} \quad \text{and} \quad E(y) = \frac{(1 - \omega)(1 - e^{-\theta})\lambda}{1 - e^{-\lambda}}.$$

While it is clear that the probability of observing positive consumption is $\text{Prob}(y > 0) = (1 - \omega)(1 - e^{-\theta})$, it is less straightforward to interpret what constitutes the probability of nonparticipation and that of a corner solution because two alternative processes, neither of which is directly linked to the positive Poisson probability model, are accounting for these probabilities. Nevertheless, from an empirical standpoint, such a hybrid model may better represent the data generating process than either model alone.

## The Double-Hurdle Model with Interdependence

The assumption that the two mechanisms are not related is not always going to be attractive, nor will it always be true. It is quite plausible that $\theta$[defined in equation (3)] can depend on the implicit travel price, $p$. Because the travel price can be solely or strongly determined by the distance-related cost component, we could easily imagine situations, for example, where an individual decides they are simply too far away from any recreation site to take trips during the season. Alternatively, if the travel price includes the opportunity cost of time, we can again see how an individual with a high opportunity cost finds recreation trip time too costly to make any trips. With these things in mind, we now drop the assumption that the two hurdles are unrelated.

If the two mechanisms generating zero outcomes under a DH specification are correlated, a bivariate double-hurdle count-data model results when the random variable $D_i$ has a discrete distribution. In the following development, the observation index $i$ has been suppressed and the reader should note that structural parameters may be parameterized to depend on $i$. The two regimes may be represented for count-data as:

(13) $$\text{Prob}(y = 0) = \text{Prob}(y^* = 0) + \sum_{k=1}^{\infty} \text{Prob}(y^* = k, D = 0),$$

and

$$\text{Prob}(y = k) = \sum_{j=1}^{\infty} \text{Prob}(y^* = k, D = j) \qquad k = 1, 2, \ldots,$$

assuming that $y = y^*$ if $y^* > 0$ and $D > 0$.

Using Holgate's bivariate Poisson model which is defined as

$$\text{Prob}(y^* = y, D = d) = \exp(-\lambda - \theta - \xi) \sum_{j=0}^{\min(y,d)} \frac{\xi^j}{j!} \frac{\lambda^{y-j}}{(y-j)!} \frac{\theta^{d-j}}{(d-j)!}$$

(Johnson and Kotz), the likelihood function of the double-hurdle count model under dependence is

(14) $$\prod_{y=0} [\exp(-\lambda - \xi) + \exp(-\theta - \xi) - \exp(-\lambda - \theta - \xi)] \cdot \prod_{y>0} \left[\frac{\exp(-\lambda - \xi)(\lambda + \xi)^y}{y!} - \frac{\exp(-\lambda - \theta - \xi)]\lambda^y}{y!}\right],$$

where $\xi$ represents the covariance between the two Poisson processes.[10] Note that (14) collapses to (10) when $\xi$ is zero.

It can be shown that

(15) $$E(y|y > 0) = \frac{\lambda(1 - \text{Prob}(D = 0)) + \xi}{\text{Prob}(y > 0)} = \frac{\lambda + \xi - \lambda e^{-\theta-\xi}}{1 - e^{-\lambda-\xi} - e^{-\theta-\xi} + e^{-\lambda-\theta-\xi}},$$

so that

(16) $$E(y) = \lambda + \xi - \lambda e^{-\theta-\xi}.$$ [11]

Under the case of an independent hurdle mechanism $\xi = 0$, expressions (15) and (16) reduce to their counterparts in expression (11). Similarly, the three categories into which survey respondents are classified suggest that the probability of nonparticipation is $e^{-\theta-\xi}$, the probability of potential participation is $e^{-\lambda-\xi}(1 - e^{-\theta})$, and the probability of reported participation is $1 - e^{-\lambda-\xi} - e^{-\theta-\xi} + e^{-\lambda-\theta-\xi}$.

---

[10] The marginal distributions of $y^*$ and $D$ are Poisson with parameters $\lambda + \xi$ and $\theta + \xi$, respectively. To derive Prob($y$ = 0) note that the second term on the RHS of (13) is

$$\sum_{k=1}^{\infty} \text{Prob}(y^* = k, D = 0) = \sum_{k=0}^{\infty} \text{Prob}(y^* = k, D = 0) - \text{Prob}(y^* = 0, D = 0)$$
$$= \text{Prob}(D = 0) - \text{Prob}(y^* = 0, D = 0).$$

So that Prob($y$=0) = Prob($y^*$ = 0) + Prob($D$ = 0) − Prob($y^*$ = 0, $D$ = 0) explaining the first expression in (14). To derive the second expression in (14), write Prob($y$ = $k$) as Prob($y^*$ = $k$) − Prob($y^*$ = $k$, $D$ = 0) since the marginal distribution of $y^*$ would result if the summation were begun at $j$ = 0.

[11]

$$E(y|y > 0) = \sum_{k=1}^{\infty} \frac{k\,\text{Prob}(y = k)}{1 - \text{Prob}(y = 0)}$$
$$= \frac{\sum_{k=1}^{\infty} k\{\text{Prob}(y^* = k) - \text{Prob}(y^* = k, D = 0)\}}{1 - \text{Prob}(y^* = 0) - \text{Prob}(D = 0) + \text{Prob}(y^* = 0, D = 0)}.$$

The numerator then may be written as $E(y^*) - \lambda\exp(-\theta - \xi) = \lambda + \xi - \lambda\exp(-\theta - \xi)$.

## Summary and Conclusions

This article has focused on count-data models of recreation demand. While multiple-site models (those that allow the demand for more than one site to be estimated as is true in a system of equations) are becoming increasingly common, with two exceptions (Shonkwiler; Ozuna and Gomez), the count-data models still are single-site models. For this class of models, we have developed the probability mass function and have derived several hurdle models that can be used to estimate the demand for recreation at one site, including a double-hurdle model which corresponds to the excess zero-modified Poisson model (Johnson and Kotz). The relationship between a double-hurdle model and some of the modified count-data models (e.g., ZAP, ZIP, ZMP) has perhaps previously not been made clear. The DH model can be interpreted in the same way that Johnson and Kotz's zero-modified Poisson model is, but it is not identical to Mullahy's single-hurdle model. Our model can be extended for use with other probability distributions.

When the population group of interest includes those who are nonusers, as well as those who are potential users, the double-hurdle model is especially attractive. For example, in an empirical application, say to water-based recreation, it would be interesting to know when some individual will be totally unresponsive to water level or other site quality changes. Such changes are not without economic implications. The individuals who respond to these sorts of changes are likely quite different from those who respond to the cost of renting a boat. Or, putting this in terms of another policy which may be controllable by agencies who manage recreation boat ramps, the potential user doesn't fish because his opportunity cost of time is too high, and the congestion at the lake increases his time cost of loading and unloading his boat. Perhaps none of this affects the nonuser.

We stress the point here that a sample which only includes users cannot be used to infer behavior for these potential users. Further, the double-hurdle models allow exploration into nonuse, as opposed to potential use. The credibility of the hurdle models is higher than the more conventional count-data models that are limited to a truncated sample of known recreators, in that the models handle the nonparticipation decision for some recreators. Our recommendation is that when modelers have the opportunity to design the survey questionnaire, as well as define the sample group, that they first consider whether some set of price/quality/quantity conditions exist which could cause potential recreators to begin taking trips. Second, we urge these modelers to design the survey questionnaire to capture nonuser's and other responses which can better explain the hurdle mechanisms in the models we have illustrated here.

In essence, we need to spend some time and energy thinking (and include some survey questions asking) about why those who could recreate outdoors do not. Some possible factors which influence the decision to stay home or take recreation trips include age, the total amount of income devoted to recreation, the number of "new" children in the home, the flexibility of work schedules, and physical ability or characteristics of the individuals. Unfortunately, while age is often asked of survey respondents, the other questions are not typical in recreation survey questionnaires. Other factors could be identified in focus groups conducted to explore recreation participation preferences and decision making. With these features of the problem being considered carefully, the best survey implementation plan will likely be much clearer to the investigators who have the luxury of designing the survey.

[*Received August 1995; final version received April 1996.*]

## References

Blundell, R., and C. Meghir. "Bivariate Alternatives to the Tobit Model." *J. Econometrics* 34(1987):179–200.

Bockstael, N., K. E. McConnell, and I. Strand. "Recreation." In *Measuring the Demand for Environmental Quality,* eds., J. B. Braden and C. D. Kolstad, pp. 227–70. North Holland: Elsevier Science Publishers, 1991.

Creel, M., and J. Loomis. "Theoretical and Empirical Advantages of Truncated Count Data Estimators." *Amer. J. Agr. Econ.* 72(1990):434–41.

Englin, J., and J. Shonkwiler. "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand under Conditions of Endogenous Stratification and Truncation." *Rev. Econ. and Statis.* 77(1995):104–12.

Greene, W. H. "Accounting for Excess Zeroes and Sample Selection in Poisson and Negative Binomial Regression Models." Unpub. draft, Department of Economics, Stern School of Business, New York University, 1994.

Grogger, J. T., and R. T. Carson. "Models for Truncated Counts." *J. Appl. Econometrics* 6(1991):225–38.

Hellerstein, D. M. "The Treatment of Nonparticipants in Travel Cost Analysis and Other Demand Models." *Water Resour. Res.* 29(1992):1999–2004.

———. "Using Count Data Models in Travel Cost Analysis with Aggregate Data." *Amer. J. Agr. Econ.* 73(1991):860–66.

Hellerstein, D. M., and R. Mendelsohn. "A Theoretical Foundation for Count Data Models." *Amer. J. Agr. Econ.* 75(1993):604–11.

Holgate, P. "Estimation for the Bivariate Poisson Distribution." *Biometrika* 51(1964):241–45.

Johnson, N., and S. Kotz. *Discrete Distributions.* New York: John Wiley, 1969.

Lambert, D. "Zero-Inflated Poisson Regression with Application to Defects in Manufacturing." *Technometrics* 34(1992):1–14.

Lin, C. T. J., and J. W. Milon. "Attribute and Safety Perceptions in a Double-Hurdle Model of Shellfish Consumption." *Amer. J. Agr. Econ.* 75(1993):724–29.

Morey, E. R., W. D. Shaw, and R. D. Rowe. "A Joint Discrete-Choice Model of Recreational Participation, Site Choice and Activity Valuation When Complete Trip Data Is Not Available." *J. Environ. Econ. and Manage.* 20(1991):181–201.

Morey, E. R., M. Watson, and R. D. Rowe. "A Repeated Nested-Logit Model of Atlantic Salmon Fishing." *Amer. J. Agr. Econ.* 75(1993):578–92.

Morey, E. R., D. Waldman, D. Assane, and W. D. Shaw. "Searching for a Model of Multiple-Site Recreation Demand That Admits Interior and Boundary Solutions." *Amer. J. Agr. Econ.* 77(1995):129–40.

Mullahy, J. "Specification and Testing of Some Modified Count Data Models." *J. Econometrics* 33(1986):341–65.

Ozuna, T., and I. A. Gomez. "Estimating a System of Recreation Demand Functions Using a Seemingly Unrelated Poisson Regression Approach." *Rev. Econ. and Statis.* 76(1994):356–60.

Parsons, G., and M. Needelman. "Site Aggregation in a Random Utility Model of Recreation." *Land Econ.* 68 (1992):418–33.

Pudney, S. *Modeling Individual Choice.* New York: Basil Blackwell, 1989.

Shaw, D. "On Site Samples' Regression: Problems of Nonnegative Integers, Truncation, and Endogenous Stratification." *J. Econometrics* 39(1988):211–23.

Shonkwiler, J. S. "Systems of Count Data Models." In *Proceedings of the W-133 Regional Research Group,* pp. 3–20. Compiled by D. Larson, Dept. of Agr. Econ., University of California, Davis, 1995.

Smith, V. K. "Taking Stock of Progress with Travel Cost Recreation Demand Models: Theory and Implementation." *Marine Resour. Econ.* 6(1989):279–310.

Yen, S. T., and W. L. Adamowicz. "Statistical Properties of Welfare Measures from Count-Data Models of Recreation Demand." *Rev. Agr. Econ.* 15(1993):203–16.