



Munich Personal RePEc Archive

# **A Theoretical Foundation for Count Data Models**

Daniel Hellerstein and Robert Mendelsohn

Economic Research Service, USDA

1993

Online at <http://mpra.ub.uni-muenchen.de/25265/>

MPRA Paper No. 25265, posted 23. September 2010 13:32 UTC

# A Theoretical Foundation for Count Data Models

Daniel Hellerstein and Robert Mendelsohn

The paper develops a theoretical foundation for using count data models in travel cost analysis. Two micro models are developed: a restricted choice model and a repeated discrete choice model. We show that both models lead to identical welfare measures.

*Key words:* count data, repeat discrete choice, travel cost analysis, welfare analysis.

For several decades, economists have used the annual demand for trips in order to measure the nonmarket value of recreation sites. Two features of trip demand functions complicate their estimation: trip demand is nonnegative and occurs in integer quantities. The fact that trip demand cannot be negative results in a censored (at zero) data set; failure to account for censoring leads to biased estimation. The integer nature of trip demand, when continuous models are estimated, can also lead to biased results.

A variety of techniques have been developed to deal with these problems, including models incorporating truncated error distributions, random utility models, discrete/continuous models, and repeated discrete choice models.<sup>1</sup> In this paper we explore the use of count data estimators, such as the Poisson model, to embody the recreational demand for trips. Poisson models are becoming increasingly common (Hellerstein, Creel, and Loomis; Smith, Shaw, Terza, and Wilson). In addition, a variety of count model extensions to the Poisson have been recently developed, providing analysts with a menu of ro-

bust and flexible estimators (see Hausman, Hall, and Griliches, or Cameron and Trivedi).

Although the attractive econometric properties of count estimators are well understood, a theoretical foundation for their use in welfare analysis has not yet been presented. In particular, the link between an individual consumer's optimization problem and a count estimator has not been drawn. Without a theoretical foundation, it is not clear how to interpret count models. More importantly, it remains ambiguous how to apply the results of count estimators of demand to welfare analysis. For example, it is unclear how to value recreation sites on the basis of a count demand model for trips.

Our paper addresses this shortcoming by developing two theoretical frameworks for count demand models. The first model modifies a standard, continuous demand model to account for a constrained integer choice set. The second approach is based on a discrete choice model which is then repeated over time. Welfare measures based on both these underlying models are derived, and are shown to yield the same formula for measuring consumer surplus.

---

Daniel Hellerstein is a natural resource economist with the Economic Research Service, Resources and Technology Division; Robert Mendelsohn is a professor at the Yale School of Forestry and Environmental Studies.

The views in this paper are those of the authors and do not necessarily reflect the views of the U.S. Department of Agriculture or Yale University.

Review coordinated by Richard Adams.

<sup>1</sup> Continuous demand models have been estimated with truncated error distributions (Shaw). Random utility models have been estimated as though the choice was purely discrete (Smith and Kaoru, Hanneman [1984a]). Continuous models have been appended to discrete models (Heckman, Bockstael et al.); with the discrete choice stage used to predict the probability of "participation," and the continuous stage used to estimate the amount of goods selected, given participation. Repeated discrete models have been explored with fixed number of choice opportunities (Morey, Shaw, and Rowe), and discrete choice has been applied given the prior number of trips taken (Adamowicz, Jennings, and Coyne).

## The Restricted Choice Model

We begin with the standard assumption that each individual maximizes utility subject to an income constraint. We add an additional constraint, however, that the choice of trips,  $X_1$ , must be a nonnegative integer. In addition, we assume that at the beginning of the season each individual chooses  $X_1$  (the number of trips) as well as the quantity of a vector of other goods ( $X_2$ ).

Following Pudney (p. 94), utility maximiza-

tion can be expressed as a function of  $X_1$  and  $\mathbf{X}_2$ , where  $\mathbf{X}_2$  is a vector of all other goods assumed to be available in any quantity. Formally, each individual solves

$$(1) \quad \max_{X_1 \in I, \mathbf{X}_2} [U(X_1, \mathbf{X}_2, \epsilon; \beta) | P * X = p_1 x_1 + \mathbf{p}_2 \mathbf{x}_2 = Y]$$

where  $\mathbf{P}$  (the vector of prices) is divided into  $P_1$  (the price of the indivisible good) and  $\mathbf{P}_2$  (a vector of prices of other goods),  $\epsilon$  are unobservable factors specific to an individual, and  $Y$  is income.<sup>2</sup> Since  $X_1$  is restricted to  $I$  ( $I = 0, \dots$ ), this can be rewritten as

$$(2) \quad \max_{X_1 \in I} \{ \max_{\mathbf{X}_2} U[(X_1, \mathbf{X}_2, \epsilon; \beta) | \mathbf{P}_2 \mathbf{X}_2 = Y - P_1 X_1] \}$$

Taking the dual of (2), the expenditure function can be written as

$$(3) \quad E[P_1, \mathbf{P}_2, \epsilon; U_0] = \min_{X_1} \{ P_1 X_1^* + [\min_{\mathbf{X}_2} (\mathbf{P}_2 \mathbf{X}_2)] \}$$

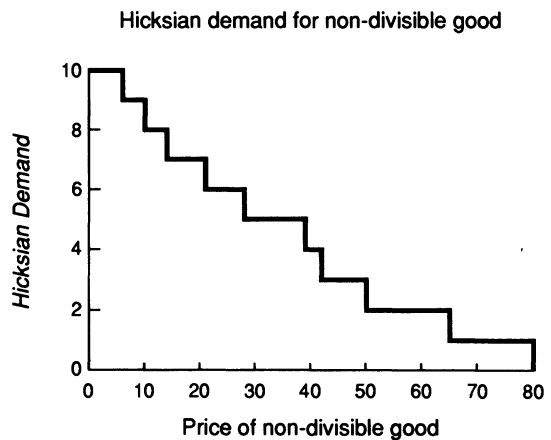
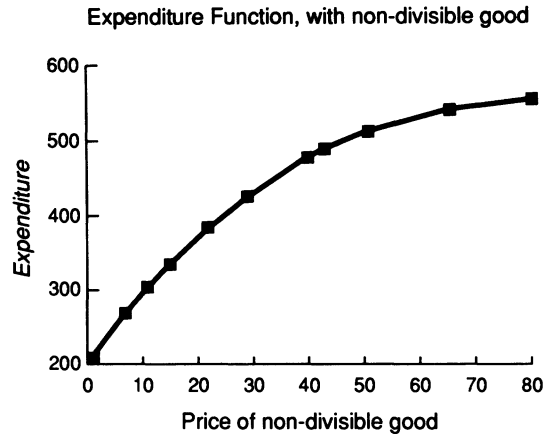
$$\text{s.t. } U(X_1^*, \mathbf{X}_2, \epsilon; \beta) = U_0, X_1^* = X_1 \}$$

where  $U_0$  is a reference level of utility.

Equation (3) highlights the two components of the decision making process: how much of  $X_1$  to consume and how much of  $\mathbf{X}_2$  to buy. Since  $X_1$  can only be changed in infra-marginal amounts, the compensated demand for  $X_1$ ,  $H(P_1, \mathbf{P}_2, \epsilon, U_0) = \partial E / \partial P_1$ , will be constant over discrete ranges of the expenditure function, with discrete jumps at prices that define the endpoints of these ranges. As illustrated in figure 1, the expenditure function will be piecewise linear, and the compensated demand will be a step function.

If repeated observations on a single individual could be obtained, each observation differing only in price, it would be possible to determine the step-function comprising the compensated demand curve for an indivisible good. In actual circumstances, such a highly controlled sample is rarely available. Instead, price variation occurs across individuals, where each individual in the sample possesses a unique set of unobservable ( $\epsilon$ ) factors. At any price, these factors (*ceteris paribus*) determine the quantity each individual consumes.

Estimating demand relationships with such data, the analyst can at best determine proba-



**Figure 1. Expenditure function, with non-divisible good. Hicksian demand for non-divisible good**

bilities of observing a level of demand, given prices, income, and other observable variables. One means of summarizing these probabilities is through a probability density function. Viewed in this manner, estimation of a demand curve is an exercise in computing the parameters of a probability density function. These parameters will vary as prices vary; hence, the probability of observing a particular level of demand will change as prices vary.

It is interesting to consider the estimation of continuous demand curves. The random component ( $\epsilon$ ) is often included as a demand shifter; for example, in the linear model of demand for a good  $Q$ ,  $Q = \mathbf{X}\beta + \epsilon$ . Alternatively, one can assume that, conditional on observed prices, demand will be distributed according to some continuous probability distribution. For example, demand can be postulated to follow a normal

<sup>2</sup> For expositional simplicity, observable individual specific factors (such as age and education) are not explicitly shown in this paper

distribution:  $Q \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2)$ ; with the  $\mathbf{X}\boldsymbol{\beta}$  of the linear model now interpreted as the location parameter of a normal distribution, and  $\sigma^2$  describing the variance of  $\epsilon$  across the population. This interpretation of continuous demand is essentially the same as the interpretation of the demand for indivisible goods offered above.

For indivisible goods, a probability distribution defined only over the nonnegative integers is required. One such candidate for this distribution is the Poisson. The Poisson probability distribution is a single parameter distribution ( $\lambda$ ), with probability density function (PDF) defined as

$$(4) \quad \text{prob}(Q = n) = \frac{e^{-\lambda} \lambda^n}{n!}; \quad n = 0, 1, \dots, \infty$$

where  $Q$  is a potential integer outcome. The  $\lambda$  parameter of the Poisson is equal to the mean,  $E[Q]$ , and the variance,  $\sigma^2[Q]$ , of  $Q$ . Typically, the  $\lambda$  parameter is modeled as a function of prices and income, such as  $\lambda(\mathbf{P}, Y; \boldsymbol{\beta})$ . For applied work, an exponential form for  $\lambda$  is usually employed: for example,  $\lambda = \exp(\beta_0 + \boldsymbol{\beta}_p \mathbf{P} + \boldsymbol{\beta}_y Y)$ , where  $\boldsymbol{\beta}$  is a vector of coefficients to be estimated.

Estimation of a Poisson model,<sup>3</sup> using data on demand for an indivisible good (such as trips to a recreational site), yields coefficient estimates which can be used to compute values of  $\lambda(\mathbf{P}, Y; \boldsymbol{\beta})$ . As a continuous quantity,  $\lambda(\mathbf{P}, Y; \boldsymbol{\beta})$  does not represent an obtainable level of demand. Rather,  $\lambda(\mathbf{P}, Y; \boldsymbol{\beta})$  parameterizes the distribution of demand (over the nonnegative integers) for individuals facing prices  $\mathbf{P}$  and income  $Y$ .

Welfare analysis is often conducted by computing a consumer surplus (as an approximation to a compensating variation) by integrating under a demand curve. With count models, the estimated function is a probability distribution of trips. Taking the expectation of this distribution yields an expected response (number of trips) at every price. By integrating underneath this expected response, a measure of the expected value of consumer surplus is obtained.

Formally, the expected value of the consumer surplus ( $E[CS]$ ), given a price change in good 1 from  $\mathbf{P}_{1a}$  to  $\mathbf{P}_{1b}$ , is

$$(5) \quad E[CS] = \int_{\mathbf{P}_{1a}}^{\mathbf{P}_{1b}} \int_E [f(\epsilon) T(p, \mathbf{P}_2, Y, \epsilon; \boldsymbol{\beta})] d\epsilon dp$$

<sup>3</sup> Estimation of a Poisson, and other count models, is usually accomplished with maximum likelihood techniques. A growing number of econometric packages directly support count model estimation; such as LIMDEP, SHAZAM, and GRBL.

where  $T(\cdot)$  is an individual's demand curve for the indivisible good (e.g.; trips), which will be a step function with exact shape dependent on  $\epsilon$ . The  $\epsilon$  argument in  $T$ , which has a range of support of  $E$  and a PDF equal to  $f(\epsilon)$ , is meant to capture the influence that unobservable factors have on trip taking decisions.<sup>4</sup> Rearranging (5) yields

$$(6) \quad \int_{\mathbf{P}_{1a}}^{\mathbf{P}_{1b}} \lambda(p, \mathbf{P}_2, Y; \boldsymbol{\beta}) dp = E[CS]$$

in which we use the assumption that trip demand is, *ceteris paribus*, Poisson distributed, and the mean of the Poisson equals  $\lambda(\mathbf{P}_1, \mathbf{P}_2, Y; \boldsymbol{\beta})$ . Note that if one estimates  $\lambda = e^{\mathbf{X}\boldsymbol{\beta}}$  ( $\mathbf{X} = \mathbf{P}, Y$ ) and  $\mathbf{P}_{1b}$  equals infinity, equation (6) yields the standard result of  $E[CS] = -\lambda/\boldsymbol{\beta}_p$ .

Summarizing this section, we show that when using count models to estimate trip demand, computation of the expectation of consumer surplus is obtained by integrating under the expected value of demand. If one assumes that demand, at any given price, follows a Poisson distribution, then the expected value of demand will equal  $\lambda$ . This result holds even though the expected value of demand may not be an integer and thus cannot be obtained by a single individual.

It is interesting to contrast this to continuous models, where consumer surplus measurement techniques dictate use of observed demand (Bockstael et al.) to estimate consumer surplus for individuals in a sample. A common presumption is that random and unobservable factors ( $\epsilon$ ) effect demand in an additive (or multiplicative) fashion. The count model, in contrast, estimates the distribution of trips from which any individual draws; with random factors incorporated in a parametric fashion rather than as a residual.

### The Repeated Discrete Choice Model

As an alternative to the restricted choice framework presented above, count models can also be derived from repeated discrete choices. At each choice interval, the consumer can make a binomial (zero/one) choice to consume or not. For example, each day, the recreator can choose whether to take a trip to a site or to engage in some other activity. The count model can then

<sup>4</sup> Note that no parametric assumptions are made about how  $\epsilon$  directly influences choices.

be derived from a repeated application of these discrete choices.

A simple conditional utility model is adopted to reflect the discrete choice of consuming or not:

$$(7) \quad V^* = \max_{j \in J} (V_j(P_j, Y, \epsilon_j); \quad J = \{0, 1\})$$

where  $V^*$  is realized utility and  $V_j(P_j, Y, \epsilon_j)$  is the utility associated with the choice of good  $j$ , given the price of obtaining activity  $j$  ( $P_j$ ), the individual's income ( $Y$ ), and a random shock term unique to good  $j$  ( $\epsilon_j$ ). Good one ( $j = 1$ , e.g. the site is visited) is selected when the choice of good one yields greater utility than realized when good zero ( $j = 0$ , e.g. not visiting the site) is chosen.

When the random shocks,  $\epsilon = \{\epsilon_0, \epsilon_1\}$ , change over time, equation 1 becomes

$$V_t^* = \max_{j \in J} (V_j(P_j, Y, \epsilon_{jt}) \quad J = \{0, 1\}.$$

Hence the consumer's choice will depend on the realization of  $\epsilon_t = \{\epsilon_{0t}, \epsilon_{1t}\}$ .<sup>5</sup> Furthermore, if we assume (without loss of generality) that  $P_0$  equals zero, on a given day  $t$  there is a probability  $\pi_t(P_1, Y)$  that the good one will be chosen (the site will be visited), and a probability  $1 - \pi_t(P_1, Y)$  that it will not.<sup>6</sup> If chosen, a quantity of one is demanded, otherwise the quantity demanded is zero.

If  $P_1$  is constant across time, and the distribution of  $\epsilon$  is independent and identically distributed (iid) across time,  $\pi_t$  will be constant over time. Therefore, the outcome of the repeat discrete choices faced by the consumer can be modeled as a series of iid draws. Total number of draws over the course of the season will have a binomial distribution. As the number of draws increases, and the probability of choice decreases proportionally, this binomial distribution will asymptotically converge to a Poisson distribution (Mood, Graybill, and Boes). In other words, the count of the number of days (within a year) that good one is chosen will be asymptotically distributed as a Poisson random variable.

It is important to note that the Poisson distribution of outcomes is not dependent on the ex-

act distribution of  $\pi$ . However, the functional form of the Poisson parameter,  $\lambda(\mathbf{P}, Y; \boldsymbol{\beta})$ , does depend on  $\pi$ . For example, it can be shown that if each discrete choice yields a logit distribution for  $\pi$ , and the choice probability is small and constant across time, the functional form for  $\lambda$  will asymptotically equal  $\exp(\beta_0 + \beta_p P + \beta_y Y)$ .<sup>7</sup>

To analyze welfare calculations in the repeat discrete choice context, the results of Hanemann 1984 (b) are adapted. We first focus on a single choice opportunity (say, a single day), and define a measure of the value of the good (say, a recreational site). For a given day  $t$ , site value is based on a compensating variation ( $CV_t$ ), defined as

$$(8a) \quad CV_t = \min_{CV} \text{ s.t. } V_1(P_0, P_1 + CV, Y, \epsilon_{1t}) \geq V_0(P_0, P_1, Y, \epsilon_{0t}).$$

Because  $\epsilon$  is stochastic,  $CV_t$  is also stochastic. Thus, it is of interest to examine the expected value of  $CV_t$ ,

$$(8b) \quad E[CV_t] = \int_0^\infty CV f_t(CV) dCV = \int_0^\infty (1 - F_t(CV)) dCV$$

where  $f_t(CV)$  and  $F_t(CV)$  are the probability density function and cumulative distribution function of  $CV_t$  (respectively).<sup>8</sup>

When  $V_0(P_0, P_1, Y, \epsilon_{0t}) \geq V_1(P_0, P_1, Y, \epsilon_{1t})$ , good one is not chosen and  $CV_t$  equals zero. Therefore

$$\begin{aligned} F_t(0) &= \text{prob}(\text{Good 1 not chosen on day } t) \\ &= 1 - \text{prob}(\text{Good 1 chosen on day } t) \\ &= 1 - \pi_t(P_1, Y), \text{ (by assumption).} \end{aligned}$$

Similarly, for any nonnegative quantity  $A$ ,  $F_t(A)$  will equal the probability of good one not being chosen when its price equals  $P_1 + A$ , which equals  $1 - \pi_t(P_1 + A, Y)$ . Substituting these results for  $F_t(0)$  and  $F_t(A)$  into (8b) yields

$$(8c) \quad E[CV_t] = \int_0^\infty \pi_t(p_1 + A, Y) dA = \int_{p_1}^\infty \pi_t(p, Y) dp.$$

<sup>5</sup> For example, if  $V_1(\cdot) \approx V_0(\cdot)$  and  $P_1 \approx P_0$ , and if on day  $t = \tau$ ,  $\epsilon_{1\tau}$  is very large and  $\epsilon_{0\tau}$  is very small, then good one will be chosen. Alternatively, if on day  $t = \gamma$ ,  $\epsilon_{1\gamma}$  is very small and  $\epsilon_{0\gamma}$  is very large, then good zero will be chosen.

<sup>6</sup> For example, if utility is of the form  $V + \epsilon$ , good one ( $j = 1$ ) is chosen when  $\epsilon_0 < \epsilon_1 + V_1 - V_0$ . Given that  $V_1$  and  $V_0$  are non-stochastic, different choices occur as  $\epsilon_1$  and  $\epsilon_0$  vary. The classic case of logit probabilities occurs when the elements of  $\epsilon_1$  and  $\epsilon_0$  are independently drawn from a type I extreme value distribution (Maddala, ch. 3).

<sup>7</sup> The logit assumes that  $V_i = \mathbf{X}\boldsymbol{\beta} + \epsilon_i$ , where  $\mathbf{X}$  is a vector of prices, etc., and  $\epsilon_i$  follows a type I extreme value distribution. See the appendix for a further discussion of these results.

<sup>8</sup> See Hanemann (1984b), equation 26, or Mood, Graybill, and Boes ch 4.1; where the assumption that  $F(CV) = 0$  for  $CV < 0$  is used. Note that  $f_t$  and  $F_t$ , which are strictly conditional on  $\epsilon$ , may be specific to "day"  $t$ .

Lastly, it is readily shown that the  $E[CV_t]$  equivalent to the change in the price of good one from  $P_a$  to  $P_b$  equals

$$(9) \quad E[CV_t] = \int_{P_a}^{P_b} \pi_t(p, Y) dp.$$

Given that  $\pi_t$  is constant over time, the expected value of the total compensating variation (CV) over an entire period (say, over a year consisting of  $T$  days) will be

$$(10) \quad E[CV] = E\left[\sum CV_t\right] = \sum E[CV_t] \\ = \sum_t \int_{P_a}^{P_b} \pi_t(p, Y) dp = \int_{P_a}^{P_b} \sum_t \pi_t(p, Y) dp$$

Because the Poisson process defines  $\lambda$  as the result of many small probability events, it immediately follows that<sup>9</sup>

$$(11) \quad E[CV] = \int_{P_a}^{P_b} \sum_t \pi_t(p, Y) dp \approx \int_{P_a}^{P_b} \lambda(p, Y) dp.$$

Therefore, the price integral over the Poisson parameter,  $\lambda(P, Y)$ , is a legitimate approximation to the compensating variation. Furthermore, as with the restricted choice model, if  $\lambda = \exp(\mathbf{X}\boldsymbol{\beta})$  is used and  $P_b = \infty$ ,  $E[CV]$  will equal  $-\lambda/\beta_p$ .<sup>10</sup>

## Extensions

The assumption that the resulting distribution of trips is Poisson need not always hold. In particular, the assumption of equality between the expected value,  $E[Q]$ , and the variance of the distribution,  $\sigma^2[Q]$ , is stringent. The Poisson relationship,  $\lambda = \lambda(\mathbf{P}, Y; \boldsymbol{\beta})$ , also does not contain an error component.<sup>11</sup>

<sup>9</sup> The Poisson parameter  $\lambda$  parameterizes the distribution of the sum of random events  $I_1 \dots I_t$ :  $\lambda = E[\sum I_t]$ , where  $I_t$  takes on values of 1 or 0, with probability  $\pi$  and  $1 - \pi$  respectively. Given that  $\pi$  is constant and  $I_t$  is independently distributed, it is readily shown that  $\lambda = \sum_t \pi_t$ .

<sup>10</sup> Note that we are approximating the CV with a Marshallian consumer surplus measure. If desired, a compensated demand curve could be estimated yielding an exact compensated Poisson welfare measure.

<sup>11</sup> Note that the lack of an error component in the estimator of  $\lambda$  implies that there are no omitted variables. In the context of the repeated discrete choice model, this implies that the daily probability of visitation can be predicted without error; it does not imply that the actual number of visits observed can be estimated exactly. The actual number of trips will be drawn from a Poisson distribution.

To relax these assumptions, a variety of extensions to the Poisson are available. An especially appealing alternative is the Negative Binomial count model. Formally (following Cameron and Trivedi), if  $Q$  is a Poisson random variable with parameter  $\lambda$ , and  $\lambda$  is distributed as a gamma random variable  $\gamma(\lambda; \mu, \nu)$ , then  $Q$  is distributed as Negative Binomial random variable with  $E[Q] = \mu$  and  $\text{var}[Q] = \mu + \mu^2/\nu$ .

The relaxation of statistical restrictions offered by the Negative Binomial can be further extended. Specifically, the Poisson (and Negative Binomial) are examples of the class of linear exponential functions, and it can be shown that as long as the specification of the mean is correct, linear exponential functions will be robust to misspecification (Gourieroux, Montfort, and Trognon). For example, as long as  $E[Q] = \lambda$ , one can consistently estimate  $\beta$  using a pseudo-maximum likelihood (PML) estimator in conjunction with the Poisson distribution (Cameron and Trivedi).

Both the restricted choice and the repeated discrete choice models are easily extended to these general count models. The restricted choice model can be described as a reduced form incorporating information on utility maximization and on unobservable factors. Therefore, use of a more sophisticated model (such as the Negative Binomial) is straightforward, and need only be justified on econometric grounds of efficiency and consistency. Earlier results on welfare calculations are also readily extended, so long as a consistent estimate of the expected value of demand is available.

For the repeat discrete choice model, it is instructive to examine the process by which a non-Poisson distribution might arise. First, consider the Negative Binomial. A gamma distribution of  $\lambda$  could arise due to variation in the underlying probability ( $\pi_t(\mathbf{P}, Y)$ ) of choosing to consume the discrete good (such as a trip to recreational site), with this probability constant across time, but varying across individuals who are otherwise similar. Knowledge of the exact distribution of the daily probability across individuals ( $\pi_{it}$ ) is unnecessary, all that is assumed is that the process gives rise to a gamma distribution of  $\lambda$ .

Considering the PML estimators, it is not necessary to assume that  $\lambda$  has a gamma distribution, all that is required is that one's model of  $E[Q]$  is correct. In the context of repeated discrete choice, this implies that the mix of daily probabilities across individuals ( $\pi_{it}$ ) that give rise to  $E[Q]$  need not be known. It is conceivable

that the  $\pi_{it}$ , the probability of visiting a site randomly, fluctuates over time.<sup>12</sup>

For the more general count models, the argument supporting the use of consumer surplus estimates as approximations to the compensating variation in the Poisson case can be readily extended. Consider the Negative Binomial. For each individual,  $\lambda$  is determined by exogenous variables and a random factor. This random factor ( $\nu$ ), which is constant over time but varies across individuals, influences the constant probability ( $\pi_{it}$ ) within each time segment. Substituting  $\pi_{it}(P_1, Y, \nu)$  into the right hand side of equation (11), the conditional expectation of  $CV$  (given individual specific factors determined by  $\nu$ ), is computed as

$$(12) \quad E[CV]|\nu = \int(\lambda(\mathbf{P}, Y)|\nu)dp.$$

The unconditional expected value of  $CV$  is then

$$(13) \quad E[CV]\nu = \int(\int(\lambda(\mathbf{P}|\nu)dp)d\nu) = \int\mu(\mathbf{P})dp$$

where  $\mu(\mathbf{P}, Y)$  is the expected value of  $\lambda$ , by assumption. In other words, a consumer surplus value obtained by integrating under  $\mu$  will approximate the  $CV$ .

Now consider the general case, where the probability of choice is not necessarily constant over time. A factor  $\nu_{it}$ , representing stochastic and systematic influences on the probability of choice at time  $t$ , is now included in  $\pi_t$ . When this probability,  $\pi_{it}(P_1, Y, \nu_{it})$ , is inserted into the right hand side of equation (11),  $\lambda$  does not result, since the Poisson assumption of iid events is no longer valid. However, if one can assume that a function,  $m(P_1, Y)$ , provides a consistent estimate of the (compensated) expected number of visits demanded,  $E[Q]$ , it immediately follows that  $\sum_t \pi_t(P_1, Y, \nu_{it})$  will equal  $m(P_1, Y)$ . Hence, we can substitute  $m(\ )$  for  $\lambda(\ )$  in the right hand side of (11).

Summarizing these results, if the probability of choice is independent and identically distributed across time and across individuals, then compensating variation is computed by integrating under the Poisson ( $\lambda$ ) parameter. If this independence does not hold across individuals, but the variation across individuals yields a gamma distribution of  $\lambda$ , then compensating variation is computed by integrating under the Negative Binomial mean ( $\mu$ ). Lastly, if all that is known is that the fluctuations in probability, both across time and across individual, yields an

$E[Y]$  that follows a known function, then compensating variation is computed by integrating under this known function ( $m$ ). Note that the latter case implies that when the Poisson model is adopted, integration under  $\lambda$  will be correct even if  $\pi_{it}$  is not iid, provided that  $E[Y]$  still equals  $\lambda$ . The key point is that one's estimator for  $E[Y]$  be correct.

## Conclusion

Count data models are an appealing tool for estimation of individual demand. This paper presents two foundations for count models: a restricted choice set and a repeat discrete choice model. All of these models generate count distributions of outcomes. The restricted choice set model presumes that the interaction between observable influences (such as price and income) and unobservable factors yields a distribution of demand that can be modeled using a count probability density function, such as the Poisson. Computing the expected value of consumer surplus is readily accomplished, assuming that one's estimate of the expected value of demand is unbiased across the relevant price range.

The repeat discrete choice model presumes that in each of many time periods an individual chooses whether or not to take a trip. If the underlying probability to take a trip is constant, the observed trip demand over a season will asymptotically follow a Poisson distribution. Other count models, such as the Negative Binomial, can be derived which permit the underlying probability (of taking a trip) to vary across otherwise similar individuals, or over the season. Welfare measures from the discrete choice model can be extended to count models.

Although the presentation of the repeat discrete choice in this paper covers several cases, a number of questions remain for future study. For example, if the proper income variable in the underlying repeat choice model is not yearly income, what value should be used? In addition, if multiple day trips and time constraints reduce the number of trips possible in a season, will the asymptotic results developed here be consistent? Lastly, how should cases be modeled when the probability of visitation later in the season depends on the realized choices made earlier?

One interesting result obtained under both models is the formula used to compute consumer surplus. This formula, which in the Poisson case equals  $-\lambda/\beta_p = -\exp(\mathbf{X}\beta)/\beta_p$ , is the

<sup>12</sup> In such cases, the conditions for a Poisson process do not hold, since  $\pi_{it}$  is not iid.

same as the standard formula used in the continuous semi-log model. Therefore, the existing count literature which has used this formula is on solid ground.

In summary, count models appear to be highly flexible tools for analyzing individual recreation data. Given their strong econometric properties and sound theoretical foundation, in many circumstances count models should become the model of choice.<sup>13</sup>

[Received December 1991; final revision received November 1992.]

## References

- Adamowicz, W., S. Jennings, and A. Coyne. "A Sequential Model of Recreation Behavior." *West. J. Agr. Econ.* 15(July 1990):91-9.
- Bockstael, N., and I. Strand, K. McConnell, and F. Arsanjani. "Sample Selection Bias in the Estimation of Recreation Demand Functions: An Application to Sportfishing." *Land Economics* 66(Feb. 1990):40-9.
- Cameron, C., and P. Trivedi. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests." *J. Appl. Econometrics* 1(January 1986):29-53.
- Creel, M., and J. Loomis. "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." *Amer. J. Agr. Econ.* 72(May 1990):434-41.
- Gourieroux, C., A. Monfort, and A. Trognon. "Pseudo Maximum Likelihood Methods: Applications." *Econometrica* 52(May 1984):701-20.
- Hanemann, M. (1984a). "Discrete/Continuous Models of Consumer Demand." *Econometrica* 52(May 1984):541-62.
- . (1984b). "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data." *Amer. J. Agr. Econ.* 66(August 1984):333-41.
- Hausman, J., B. Hall, and Z. Griliches. "Econometric Models for Count Data with an Application to the R&D Relationship." *Econometrica* 52(July 1984):909-37.
- Heckman, J. "Sample Selection Bias as a Specification Error." *Econometrica* 47(January 1979):153-61.
- Hellerstein, D. "Using Count Data Models in Travel Cost Analysis With Aggregate Data." *Amer. J. Agr. Econ.* 73(August 1991):860-66.
- . "GRBL: An Econometric Package For Count and Other Models." Available from D. Hellerstein, ERS/RTD, 1301 New York Ave NW, rm. 438, Washington DC 20005. 1992.
- Kling, C. "Estimating the Precision of Welfare Estimates." *J. Environ. Econ. and Manag.* 21(November 1991):244-59.
- Manski, C., and D. McFadden. *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge: MIT Press, 1981.
- Mood, A., F. Graybill, and D. Boes. *Introduction to the Theory of Statistics*. New York: McGraw Hill Publishing, 1974.
- Morey, E., W. D. Shaw, and R. Rowe. "A Discrete-Choice Model of Recreational Participation, Site Choice, and Activity Valuation When Complete Trip Data are Not Available." *J. Environ. Econ. and Manag.* 20(March 1991):181-201.
- Mullahy, J. "Specification and Testing of Some Modified Count Models." *J. Econometrics* 33(December 1986):341-65.
- Pudney, S. *Modeling Individual Choice: The Econometrics of Corners, Kinks and Holes*. Basil Blackwell: New York, 1989.
- Shaw, D. G. "On-Site Samples' Regression: Problems of Non-negative Integers, Truncation, and Endogenous Stratification." *J. Econometrics* 37(Feb. 1988):211-23.
- Smith, V. K. "Selection and Recreation Demand." *Amer. J. Agr. Econ.* 70(Jan. 1988):29-36.
- Smith, V. K., and Y. Kaoru. "Modeling Recreation Demand Within a Random Utility Framework." *Economic Letters* 22(1986):295-301.
- Small, K., and H. Rosen. "Applied Welfare Economics with Discrete Choice Models." *Econometrica* 49(Jan. 1981):105-30.
- Terza, J., and P. Wilson. "Analyzing Frequencies of Several Types of Events: A Mixed Multinomial-Poisson Approach." *Rev. Econ. and Statist.* 72(February 1990):108-15.
- Willig, R. "Consumer Surplus Without Apology." *Amer. Econ. Rev.* 66(Sept. 1976):589-97.

## Appendix

### Derivation of $\lambda$ from a Logit Model of Discrete Choice

To illustrate the equivalence of count models and the repeat discrete choice model, a Monte Carlo analysis of a repeated discrete choice model is performed. We start with a known random utility model, defined over the decision of whether or not to visit a site. The repeated discrete choice model is formed by generating  $T$  choices from the known random utility model, with each choice dependent on the realization of a random shock. The total number of visits, in this  $T$  day "season," is simply the number of times that the decision is made to visit the site. In addition, given knowledge of the random utility model, the welfare implied by these decisions is easily computed, and can be expressed as a compensating variation.

To start, we assume that the random utility model has the following form:

$$V_j = W_j + \epsilon_j;$$

with  $W_j = B_{aj} + B_y(Y - P_j)$ ,  $\epsilon_j$  an independently and iden-

<sup>13</sup> Furthermore, extension of these results to include site attributes and multiple sites is straightforward (an appendix discussing such extensions is available from the authors).



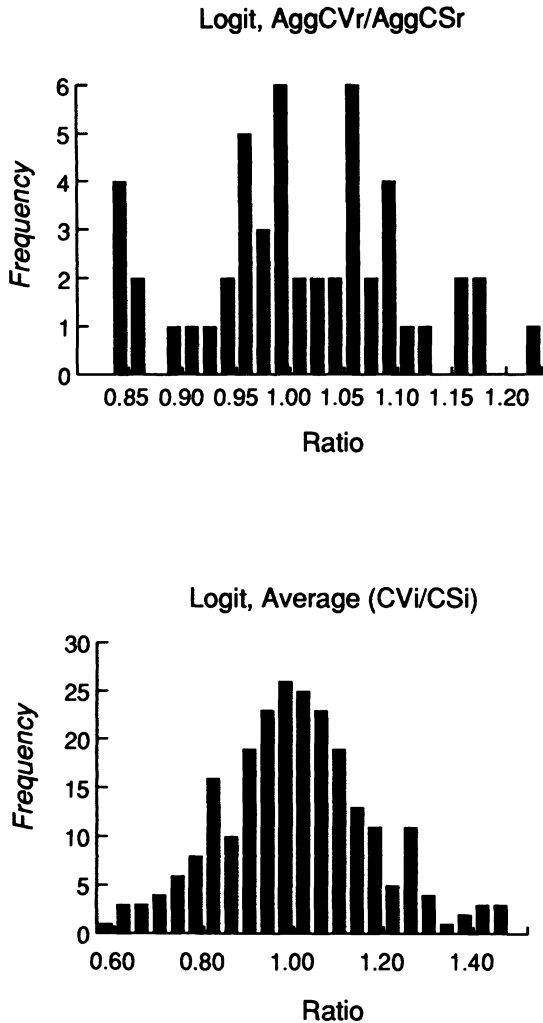


Figure 2. Logit, AggCVr/AggCSr. Logit, Average (CV<sub>i</sub>/CS<sub>i</sub>)

tically distributed Type I Extreme Value random variable,  $j = 1$  for visit and  $j = 0$  for not visit. The individual chooses to visit the site when the utility from visiting the site ( $V_i$ ) exceeds the utility from not visiting ( $V_0$ ). The price of the activity reduces residual income ( $Y - P_i$ ), for which constant marginal utility ( $B_y$ ) is assumed. Given the assumption on the error distribution, it is well known that the probability of visiting (choosing  $j = 1$ ) follows the logit form, with

$$(i) \quad \text{prob}(\text{Visit}) = \frac{\exp(W_i)}{\exp(W_i) + \exp(W_0)} = \frac{\exp(B_{a1} - B_y(Y - P_i))}{\exp(B_{a1} - B_y(Y - P_i)) + K}$$

where  $K = \exp(B_{a0} + B_y(Y - P_0))$

If the probability of choosing to visit the site is small, then  $\exp(W_i)$  will be much smaller than  $K$ . Thus, the denominator of (i),  $\exp(W_i) + K$ , can be approximated by  $K$ , so that  $\text{prob}(\text{Visit}) = \exp(W_i)/K$ . Assuming (without loss of generality) that  $P_0$  equals zero, under the Poisson model with  $\lambda = \exp(\beta_0 + \beta_p P + \beta_y Y)$ , the repeat discrete choice story requires that  $\pi_i = E[Y]/T = \lambda/T$ . Equating these two probabilities,  $\pi_i = \lambda/T$  and  $\text{prob}(\text{Visit}) = \exp(W_i)/K$ , yields

$$e^{\beta_0 + \beta_y Y + \beta_p P} \approx e^{B_{a1} + \ln\left(\frac{T}{K}\right) + B_y(Y - P)}$$

In other words, the net result of the repeated discrete choice process (in terms of total visits made) can be approximated as Poisson with  $\lambda = \exp(\mathbf{X}\boldsymbol{\beta}) = \exp(\beta_0 + \beta_y Y + \beta_p P)$ .<sup>14</sup> Moreover, the price coefficient from the count model ( $\beta_p$ ) approximates the price responsiveness coefficient in  $W_i$  ( $B_p$ ), and the constant term from the count model,  $\beta_0$ , is a reduced form incorporating  $B_{a1}$ ,  $T$  and  $K$ . Lastly, the consumer surplus generated by estimating a Poisson model and using the resulting coefficients to compute the CS will be an accurate measure of the true CV.

The accuracy of the Poisson model, when a logit repeat discrete choice process is generating the data, is examined using a Monte Carlo analysis (the details of the Monte-Carlo analysis are available from the authors upon request). Briefly, a large number of individuals are created, and for each individual a long (large  $T$ ) repeated discrete choice process is generated. CV measures are computed, as well as the number of visits per individual. The number of visits are used in a Poisson model, and the results of the Poisson model are used to form consumer surplus estimates.

This process is repeated 50 times, with the results displayed in figure 2. For each model, the frequencies of two ratios are displayed: aggregate CS over aggregate CV, and average CS over average CV.<sup>15</sup> The results clearly show that CS measures from the count model closely approximate the underlying CV (from the repeat discrete choices), with the average ratio of CV/CS quite close to 1.0.<sup>16</sup>

<sup>14</sup> Note that the approximation becomes more exact as  $\exp(W_i) + K$  approaches  $K$ .

<sup>15</sup> Aggregate CS is computed as  $\sum CS_n$ , with the sum over all  $n = 1, \dots, N$  individuals and with  $CS_n$  computed at individual  $n$ 's own price. Aggregate CV is computed as  $\sum CV_n$ . Averages of an individual's CS and CV are taken over the 50 iterations.

<sup>16</sup> Averaging over all 50 replications:  $E[\text{Aggregate CV}/\text{Aggregate}(CS)] = 1.012$  (the 90% empirical confidence interval is between 0.847 and .21). Averaging over 250 individuals:  $E[\text{Average CV}/\text{Average CS}] = 1.002$  (the 90% empirical confidence interval is between 0.662 and 1.29).