Demand and welfare effects in recreational travel models
a bivariate count data approach

Hellström, Jörgen; Nordström, Leif Jonas

Publication date:
2009

Document version
Publisher's PDF, also known as Version of record

Demand and Welfare Effects in Recreational Travel Models: A Bivariate Count Data Approach*

Jörgen Hellström[a] and Jonas Nordström[a,b,c]

[a]Department of Economics, Umeå University, SE-901 87 Umeå, Sweden

[b]Department of Economics, Lund University, Box 7082, SE-220 07 Lund, Sweden

[c]Institute of Food and Resourse Economics, University of Copenhagen, Rolighedsvej 25, 1958 Frederiksberg C, Denmark

*Corresponding author Jonas Nordström, Department of Economics, Lund University, Box 7082, SE-220 07 Lund, Sweden. e-mail: Jonas Nordstrom@nek.lu.se, Phone: +46 46 222 4671, Fax: +46 46 222 26 13. The financial support of the Swedish Research Council for Environmental, Agricultural Sciences and Spatrial Planing and the Wallander Foundation is gratefully acknowledged.

Demand and Welfare Effects in Recreational Travel Models: A Bivariate Count Data Approach

In this paper we present a non-linear demand system for households' joint choice of number of trips and days to spend at a destination. The approach, which facilitates welfare analysis of exogenous policy and price changes, is used empirically to study the effects of an increased CO_2 tax. In the empirical study, a bivariate zero-inflated Poisson lognormal regression model is introduced in order to accommodate the large number of zeroes in the sample. The welfare analysis reveals that the equivalent variation (EV) measure, for the count data demand system, can be seen as an upper bound for the households welfare loss. Approximating the welfare loss by the change in consumer surplus, accounting for the positive effect from longer stays, imposes a lower bound on the households welfare loss. From a distributional point of view, the results reveal that the CO_2 tax reform is regressive, in the sense that low income households carry a larger part of the tax burden.

Key Words: demand analysis, welfare effects, CO_2 tax, count data, bivariate zero inflation.

1 Introduction

In this paper we empirically evaluate and analyze welfare effects and changes in recreational demand due to increases in environmental taxes. More specifically, we examine the effect of an increased carbon dioxide tax, which aims to reduce the emissions of CO_2 and other greenhouse gases. The modeling approach considered in this paper accommodates for the count data feature of recreational demand, i.e., the number of trips and the number of days stayed, and treats the households' decision of number of trips and number of days to stay as a simultaneous choice. The approach renders a non-linear recreational demand system, which is used to calculate exact as well as approximative welfare measures, including/not including the welfare change due to changes in the length of the trips. The evaluation of demand and welfare effects relating to recreational activity is likely to be important in the future since many countries are committed to reducing the emissions of CO_2 and other greenhouse gases, at the same time as households' budget shares for recreational services can be assumed to be increasing with rising incomes and more leisure time.

According to the Kyoto Protocol, the overall emissions of greenhouse gases from developed countries should be at least 5 percent below 1990 levels in the commitment period 2008–2012. The commitment by the European Union (EU) is for an 8 percent reduction for the same period. Additionally, a few countries (e.g., Germany, Sweden, and the UK) have adopted a more ambitious environmental policy than required by international agreements. The UK, for example, has a national reduction goal of 20 percent for CO_2 emissions. Some US states (e.g., California, Florida, New Jersey, New York and Pennsylvania) have also adopted an environmental policy that aims to reduce the emissions of carbon dioxide, although the US has not yet ratified the Kyoto Protocol.

In Sweden the transport sector accounts for roughly 40 percent of the emissions of carbon dioxide. Two-thirds of these emissions derive from passenger transport. Thus, it is in the transportation sphere that one can expect to find the greatest potential for emission reductions by households in the future. Higher taxes on passenger transport will not only have welfare implications for the household sector, but will also affect other sectors in the economy, such as the tourism and leisure industry. These effects depend to a large extent on how price sensitive households are, and on the substitutions between the number of trips and days on vacation. Previous studies that have considered welfare measurement in recreational count data demand systems (e.g., [22] and [7]) have not considered duration of stay as an endogenous variable. In this paper, we provide some empirical results concerning different ways of measuring household welfare effects, when the household make simultaneous choice of number of trips and days to stay.

Modern recreational demand modeling usually utilizes some type of count data model to accommodate the integer-valued nature of the household's recreational demand, usually measured in terms of the number of trips. A number of authors have also considered time on site (the number of days/nights) as endogenously determined, e.g., [18], [17], [2], [11]. In the present paper both of these features are accommodated. A non-linear (Poisson) demand system is specified and used to derive appropriate welfare measures. In contrast to most earlier empirical studies, the paper considers simultaneous estimation of the demand for trips and days in a count data regression framework. Since the data have an excess amount of zeros (see e.g., [15]), i.e., there is a large probability mass at zero not consistent with most conventional count data distributions (e.g., Poisson, negative binomial), a bivariate zero-inflated Poisson lognormal (BZIPLN) model is introduced.[1] Advantages with the BZIPLN model are that compared to similar hurdle specifications, see [12], the likelihood function is relatively simpler facilitating estimation. In addition count data models with lognormal mixture densities frequently provide better fit to data [26]. A further advantage with the chosen specification is that the Poisson lognormal distribution does not constrain the correlation between the two endogenous variables to be positive (as in most other count data models, see for example [20]).[2] The paper can be viewed both as input to the evaluation of the effects and costs of Sweden's environmental policy and as input on future policy recommendations.

The outline of the paper is as follows: Section 2 presents the economic framework and introduces the empirical study. In Section 3 the data are presented and discussed. Section 4 discusses the econometric model specification and estimation, and Section 5 presents the empirical results. The concluding section contains a number of final observations.

2 The Economic Structure

In the modeling of recreational demand a number of different approaches have been used. The literature includes among other things models that consider the discrete choice of which sites to visit (e.g., [19] and [24]) and studies that focus on the number of trips a persons undertakes (e.g., [15] and [8]). To account for differences in the length of the stay, the approach has been to estimate different models depending on the duration of the trip. From both a demand and a welfare economic point of view, it is of interest to consider models that can accommodate the duration of the stay in a more flexible manner.

In this study we allow time on site to be endogenous and consider the choice of the number

[1] The bivariate zero-inflated negative binomial model was studied by [25].

[2] [23] was the first to apply the lognormal Poisson model to an economic problem.

of trips (x_1) and the total number of days to stay on the trips (x_2) a simultaneous decision.[3] Earlier studies that have treated time on site as endogenous are, for example, [17] and [12]. In the modeling the recreational choice is considered as a short-run decision conditioned on longer-run labor supply (l). As we do not want to place any restrictions on the individual's attitude to work, labor supply is included as a conditional good in the optimization problem (the most common assumption in the literature has been that the marginal utility of work time is zero, thereby linking the value of time to the wage rate). A nice feature of this approach is that consistency with microeconomic theory does not hinge at all on whether the individual is at a corner solution in the labor/leisure choice or not [3].

Due to data limitations it is not possible to observe the household consumption of other goods $\mathbf{w} = (w_1, ..., w_r)$. However, through the budget identity $y - \mathbf{p}'\mathbf{x} \equiv \mathbf{q}'\mathbf{w} \equiv m$, where $\mathbf{p}$ and $\mathbf{q}$ are prices for the goods in $\mathbf{x}$ and $\mathbf{w}$ and y is the household's total income, total expenditures on $\mathbf{w}$ are observed, i.e., m. This implies that the demand for trips and days can be specified as an incomplete demand system, see e.g., [16], [9] and [10]. Conditional on labor supply and household characteristics ($\mathbf{k}$), the conditional quasi-utility function associated with the incomplete demand system can be represented by

$$u = (x_1, x_2, m, \mathbf{q}; l, \mathbf{k}).$$

Besides the usual properties of a utility function for fixed $\mathbf{q}$ (quasi-concave, twice differentiable) this utility function possesses the properties of joint weak complementarity [21], i.e., $\partial u(0, x_2, m, \mathbf{q}; l, \mathbf{k})/\partial x_2 = 0$ and $\partial u(x_1, 0, m, \mathbf{q}; l, \mathbf{k})/\partial x_1 = 0$. This makes it acceptable to assume an interior solution, see e.g., [18]. This approach, where the individual chooses the total number of days, implies that total time is valued; but how total time is packaged into shorter or longer stays on site is a matter of indifference to the individual, aside from the effects on more or less travel time and increased or decreased travel costs. The maximization of the utility function is done subject to the budget constraint $\sum_{i=1}^{2} p_i x_i + m \leq y$, where p_1 is the travel cost per trip and p_2 is the cost per day on site. The price per day on site includes expenditures on accommodation, restaurants, shopping, activities, and on site travel. This information is available in the TDB. The observed market demands for trips and days will then be given by the function

$$\mathbf{x} = f(\mathbf{p}, \mathbf{q}, y; l, \mathbf{k}).$$

The count data structure of the dependent variables makes us assume that they have an

[3] With the individual choosing the total number of trips and the total days to stay the model is linear in the constraints. If the model were set up so the individual chose on-site time and trips, the model would be non-linear in the constraints. The choice of on-site time would affect the price of a trip. The properties if this model is outlined in [17].

exponential mean function to ensure a non-negative estimate of the number of trips and total number of days to stay. The observed demand functions for a household can thus be expected to have the form

$$x_i = \exp\left(\alpha_i(q,\mathbf{k}) + \sum_{j=1}^{2}\beta_{ij}p_j + \gamma_i y + \delta_i l\right), \qquad i = 1, 2, \tag{1}$$

where the α function is a demand shifter which depends on household characteristics. Since all prices and income are assumed to have been deflated by a linear homogeneous function of the prices for $\mathbf{w}$, the demands are zero degree homogeneous in prices and income. As income is greater than total expenditures on recreation, there is no adding-up restriction. Therefore, to have an integrable demand system, the only equality constraint is the symmetry of the Slutsky substitution terms $s_{ij} = \partial x_i/\partial p_j + x_j \partial x_i/\partial y$, i.e.,

$$\beta_{ij}x_i + \gamma_i x_i x_j = \beta_{ji}x_j + \gamma_j x_j x_i.$$

One set of restrictions consistent with this requirement is $\gamma_i = \gamma_j$ and $\beta_{ij} = \beta_{ji} = 0$.[4] Although the restrictions imposed on the demand system appear severe, the requirement of zero cross-price effects are largely unavoidable when adapting an integrability consistent Poisson demand system. Note, however, that the compensated cross-price effect between trips and day to stay might be non-zero. The expression for the compensated cross-price effect is calculated from the Slutsky equation as $s_{ij} = x_i(\partial x_j/\partial m) = \gamma x_i x_j$, [7].

The quasi-indirect utility function associated with the restricted demand functions is

$$v(\mathbf{p}, y; l, \mathbf{k}) = -\frac{\exp(-\gamma y)}{\gamma} - \sum_{i=1}^{2}\frac{\exp(\alpha_i + \beta_{ii}p_i + \delta_i l)}{\beta_{ii}}, \qquad \gamma > 0 \tag{2}$$

and is used in the calculations of Hicks' (1942) measure of equivalent variation (EV). For a price change from p^0 to p^c, EV can be written as

$$EV = -\frac{1}{\gamma}\ln\left[\exp(-\gamma y) + \gamma\left(\frac{\exp(\alpha_1 + \beta_{11}p_1^c + \delta_1 l)}{\beta_{11}} - \frac{\exp(\alpha_1 + \beta_{11}p_1^0 + \delta_1 l)}{\beta_{11}}\right)\right] - y, \tag{3}$$

for a positive income effect, $\gamma > 0$.

[4] Another set of possible restrictions would be $\alpha_i = (\beta_{ii}/\beta_{jj})\alpha_j > 0$, $\gamma_i = \gamma_j$, and $\beta_{ik} = \beta_{jk} = \beta_{kk}\forall k$. With a negative own price effect this restriction would imply that trips and numbers of days to stay are forced to be complements, and not substitutes as the empirical analysis shows. Empirically [6] impose both sets of possible restrictions, while [22] did not use the appropriate restrictions in their application.

Since the EV measure neglects the substitution possibility to longer stays, we will also estimate a model without any parameter restrictions and use the change in consumer surplus (ΔCS) as an approximate welfare measure. The change in consumer surplus due to an increased CO_2 tax may be written as

$$\begin{aligned} \triangle CS &= \int_{p_1^0}^{p_1^c} \exp(\boldsymbol{\alpha}_1 + \beta_{11}p_1 + \beta_{12}p_2 + \gamma_1 y + \delta_1 l)dp_1 - \\ &\quad \int_{p_1^0}^{p_1^c} \exp(\boldsymbol{\alpha}_2 + \beta_{12}p_1 + \beta_{22}p_2 + \gamma_2 y + \delta_2 l)dp_1 \\ &= \frac{1}{\beta_{11}}[\exp(\boldsymbol{\alpha}_1 + \beta_{11}p_1^c + \beta_{12}p_2 + \gamma_1 y + \delta_1 l) - \exp(\boldsymbol{\alpha}_1 + \beta_{11}p_1^0 + \beta_{12}p_2 + \gamma_1 y + \delta_1 l)] - \\ &\quad \frac{1}{\beta_{21}}[\exp(\boldsymbol{\alpha}_2 + \beta_{12}p_1^c + \beta_{12}p_2 + \gamma_2 y + \delta_2 l) - \exp(\boldsymbol{\alpha}_2 + \beta_{12}p_1^0 + \beta_{12}p_2 + \gamma_2 y + \delta_2 l)]. \end{aligned} \quad (4)$$

for a positive substitution effect. Although EV can usually be considered as an exact welfare measure, in our count data demand system it can be seen as an upper (lower) bound of the welfare loss since it does not account for the positive (negative) substitution effect concerning the number of days to stay.

3 Data

To estimate the model we use monthly data obtained from the Tourism and Travel Database (TDB) which covers the period January 1990 to August 1996. The TDB is a monthly telephone survey covering the population of Swedish households aged 0-74 years. Approximately 28 000 people are interviewed each year using a computer-assisted telephone interviewing technique. The TDB classifies trips as either mainly for business or for recreation. Since the interest of the paper concerns household welfare effects, the empirical study is limited to recreational trips. The survey contains, among other things, information on the number of overnight trips made during the previous month, as well as socioeconomic information. For the two most recent trips, detailed information is available on for instance the origin and destination of the trip, the main purpose of the trip, and expenditure at the destination.

The sample used in the study has been obtained after a number of restrictions on the basic data set. Households with a total number of nights greater than 30 per month and an income over SEK 800 000 were deleted from the sample, to avoid extreme values in the sample. By imposing the income restriction the sample was reduced by 0.3 percent, the mean income amounts to SEK 243 000. As we have to estimate the transport cost, we also excluded households with individuals over 65 years, since this visitor group is able to travel at a reduced rate by public transport, which is difficult to capture in practice.

In order to speed up the estimation time, that is rather long due to the large amount of variables and the numerical integration procedure used in the estimation, the final sample has been randomly sampled (approximately 20 percent of the observations for each year) from the restricted larger sample. The means of the variables in the final sample are close to the means in the larger sample and estimation of a reduced model on both the larger and the final sample indicate a high correspondence. Table 1 shows the distribution of trips and days for the 19 726 observations in the final sample.

[Table 1 about here]

Since the structure of the TDB survey limits the information available concerning trip details (as the number of days on a trip) to the two latest trips the observed total number of days during a month is censored for households making more than two trips. Thus, for a household with 4 trips the observed number of days is (based on two out of these trips), for example, 5 or more. Since only 4 percent of the households make more than two trips this feature of the data is ignored in the empirical study. [13] study the effect of accounting for this feature (endogenous censoring) using similar data and find that the effect of not accounting for censoring for this data is ignorable. Conditional on trip participation, the mean number of trips and days are 1.57 (s.e. 1.28) and 4.40 (s.e. 4.13), respectively.

3.1 Variables

The theoretical model specifies a number of variables to include in the demand system. Some are directly observable in the TDB, such as the price or cost at the destination, whereas others are indirectly observable.

A drawback with the TDB is that the total cost of transportation is not reported. Therefore, the transportation cost is calculated based on the reported origin and destination of a trip. The transportation costs are calculated for the full household. For travel by car, distance traveled is used to compute the cost. It is assumed that decision makers only consider direct costs, i.e., gas. We used the average monthly gas prices during each year from 1990-1996. Gas prices from 1990 were used together with a gas price index to calculate gas prices for other periods. Data on fuel consumption per kilometer were obtained from the Swedish Automobile Association for each year. Bus costs are calculated using a ticket price per km obtained from bus price schedules, in combination with the distance travelled. For air transportation, costs are calculated using price schedules and timetables obtained from SAS (Scandinavian Airline Systems). Air costs are based on the price for the summer of 1994 and the prices for the other periods are obtained using a monthly price index for domestic flights. Households are assumed to have used the closest airport to the reported origin of their trip. Based on household characteristics, the number of adults and children in different ages, seven different combinations of air fares are possible at each airport. Train costs are calculated using an average fare price per km obtained from Swedish Railways. We assumed that travelers who travel more than 600 km purchase a sleeper ticket, with a price corresponding to an average of the price in compartments with three and six beds. The prices are based on actual fares received by the operator, i.e., discounts are accounted for. For households with zero trips, we predict the market prices for transport and the prices at the destination by a linear model based on household characteristics.

Variables containing socioeconomic information are also used in the study. To control for possible age effects, a variable (age) containing the age of the oldest household member is used. Variables for the number of adults in the household and the number of children aged 0–6, 7–12, and 13–18 are also constructed to control for household composition effects. A dummy for the month of July is included to account for the main holiday season. Variables to control for different purposes of the trips are also included. The most common reported purposes of travel are visiting relatives and friends and visiting vacation homes. Since it is possible that households with these purposes may behave differently, e.g., the price at location may be close to zero, dummies are included for households with these reported purposes. The dummies are one if the purpose is visiting relatives and friends and vacation homes, otherwise zero. The reported purpose of the household's first trip is used as a proxy for the second trip.

The information in the TDB concerning labor supply is restricted to terms of employment for one of the adults in the household. Therefore, to account for labor supply, we include dummy variables for different terms of employment, such as part-time worker and full-time worker. Although we cannot observe the exact number of hours worked, the dummy variables will capture the main properties of labor supply that are of interest in a model for leisure days–that is, we will capture the time constraints that different terms of employment place on leisure day demand. For example, one can expect that full-time workers will usually demand at most two guest nights per week. Destination dummy variables are also added to the empirical model and works as demand shifters. Table 2 gives descriptive statistics for the explanatory variables.

[Table 2 about here]

4 The Econometric Model

To empirically model the demand for trips, x_{1h}, and the total demand for days to stay on these trips, x_{2h}, for household h, a bivariate count data regression model is specified. To account for possibly negatively correlated count variables a bivariate Poisson lognormal model is chosen. Since there are a large amount of zero observations in the sample, the model is extended to accommodate for this feature of the data. This is accomplished by inflation of the "zero-zero" probability. Since the data only includes trips with a positive number of nights, i.e., a trip is only recorded if there is a positive number of nights, it is not possible to observe the outcome one trip-zero nights. Hence, the structure of the data is either $(x_{1h} = 0, x_{2h} = 0)$ or $(x_{1h} > 0, x_{2h} > 0)$.

Assume that the total number of trips and the total number of nights have independent Poisson distributions conditional on random unobserved heterogeneity components ε_{1h} and ε_{2h} and explanatory variables z_{1h} and z_{2h}:

$$x_{ih}|\mathbf{z}_{ih}, \varepsilon_{ih} \sim P(\mu_{ih}),\ i = 1,\ 2$$

where the mean parameters (the demand functions) are specified as $\mu_{ih} = \exp(z'_{ih}\beta_i + \varepsilon_{ih}) \geq 0$ and the unobservable variables ε_{ih} are assumed to be jointly normally distributed, i.e.,

$$(\varepsilon_{1h}, \varepsilon_{2h}) \sim N\{(0,0), (1, \rho\sigma_2, \sigma_2^2)\}, \qquad |\rho| \in [0,1]$$

with σ_1^2 normalized to 1 to simplify estimation. The sign of the correlation between trips and nights is determined by the sign of ρ (the correlation between the unobserved heterogeneity terms) which is allowed to be negative.

The bivariate zero-inflated Poisson lognormal (BZIPLN) model is specified as

$$\Pr[x_{1h}{=}0, x_{2h}{=}0] = \pi_h + (1-\pi_h)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mu_{1h}) \times \exp(-\mu_{2h}) f(\varepsilon_{1h},\varepsilon_{2h})\, d\varepsilon_{1h} d\varepsilon_{2h}, \quad (5)$$

$$\Pr[x_{1h}{>}0, x_{2h}{>}0] = (1-\pi_h)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(-\mu_{1h})\mu_{1h}^{x_{1h}}}{x_{1h}!}\frac{\exp(-\mu_{2h})\mu_{2h}^{x_{2h}}}{x_{2h}!} f(\varepsilon_{1h},\varepsilon_{2h})\, d\varepsilon_{1h} d\varepsilon_{2h}, \quad (6)$$

where $\pi_h = \exp(\widehat{\mathbf{z}_h}'\boldsymbol{\theta})/(1 + \exp(\widehat{\mathbf{z}_h}'\boldsymbol{\theta})) \geq 0$, is the "inflation" parameter, parameterized as a function of the observable vector of covariates $\widehat{\mathbf{z}_h}$ and the parameter vector $\boldsymbol{\theta}$. To ensure that $\pi_h \in [0,1]$, a logistic function is utilized for π_h. The joint log-likelihood function is given by

$$l = \sum_{h=1}^{H}(1-d_h)\ln(\Pr[x_{1h}{=}0, x_{2h}{=}0]) + d_h \ln(\Pr[x_{1h}{>}0, x_{2h}{>}0]),$$

where d_h is an indicator variable that takes the value 1 if $(x_{1h}>0, x_{2h}>0)$ and 0 otherwise.

A closed form for the BZIPLN mixture is not available and Gauss-Hermite quadrature is therefore utilized to evaluate the integrals (equations 5 and 6). A one-dimensional integral can be obtained by factorization of $f(\varepsilon_1\varepsilon_2)$ into a conditional and a marginal distribution. Details concerning the Gauss-Hermite quadrature are given in Appendix A. Estimation by simulated maximum likelihood (SML) for the basic type of the bivariate Poisson log-normal model has previously been studied by [20]. [5] use Markov Chain Monte Carlo methods for the same model while [12] utilizes SML estimation for a truncated version of the model.

5 Estimation results

The estimation result for the restricted BZIPLN model is presented in Table 3. The own price coefficients for both the trip and day equations are significantly negative, with the day equation more price sensitive than the trip equation. The mean price elasticities are calculated as

$$e_{ij} = \frac{1}{H}\sum_{h=1}^{H}\frac{\partial E[x_{ijh}|\mathbf{z}]}{\partial p_{ijh}}\frac{p_{ijh}}{E[x_{ijh}|\mathbf{z}]} = \frac{1}{H}\sum_{h=1}^{H}\beta_{ij}p_{ijh}, \quad i,j \in 1,2,$$

where $\partial E[x_{ijh}|\mathbf{z}]/\partial p_{ijh} = \beta_{ij}E[x_{ijh}|\mathbf{z}]$, which gives the mean own price elasticity $e_{11} = -0.24$ for the number of days and $e_{22} = -0.13$ for the number of trips. The estimated price coefficient in the π function shows the expected sign, as a higher price reduces the probability that a household will undertake a trip, i.e., a higher price increases the probability of observing an $(0,0)$ outcome. The table also reveals a positive income effect for trips and days, although this is insignificant. The significant income effect in the π also increases the demand, as a higher income will reduce the probability that a household will stay at home.

The effects from the labor supply variables are generally insignificant in the number of trips and the π equation. However, the lengths of the stays are significantly affected by the household's labor supply. Thus, the result indicates that the number of trips is separable from labor supply while the demand for number of days is not. In relation to full-time working households, the results indicate that households classified as part time-workers, students, or home workers will generally stay for a longer time. Since full-time workers usually undertake their leisure trips at weekends, with at most two days per trip, these results seem reasonable.

The presence of children in the household will generally reduce the number of trips and prolong the length of visits, although the effects are only significant in the trip equation. The variables representing visits to vacation homes, friends/family, and the July dummy are generally significant and increase both the number of trips and the number of days. The estimated correlation coefficient ρ is positive and significant, indicating positive correlated unobserved heterogeneity.

The estimation results from the unrestricted model specification, reported in Table 4, are relatively robust compared to the results from the restricted model. The cross-price effects are, however, significant in both equations, with a positive substitution effect from a higher transportation price on the demand for days. The estimated cross-price elasticity for trips is, $e_{12} = -0.58$, while it amounts to $e_{21} = 0.33$ for the number of days.

By including the cross prices, the estimated own price coefficient in the trip equation decreases from -0.034 (s.e. 0.016) to -0.016 (s.e. 0.016), whereas we obtain an increase in the own price sensitivity in the day equation. The estimated mean own price elasticities for trips and days are -0.06 and -0.41 respectively in the unconstrained model.

By removing the parameter restriction, $\gamma_i = \gamma_j$, the unrestricted model also reveals significantly positive income effects for both the trip equation ($\gamma_1 = 0.042$ with s.e. 0.015) and the day equation ($\gamma_2 = 0.027$, s.e. 0.015).

[Table 3 about here]

[Table 4 about here]

5.1 Welfare effects

In the calculations of welfare effects, a scenario is considered where the CO_2 tax is increased by 100-percent. In the simulation we use the baseline taxes for 1998. In this year the excise duty, measured as the share of the producer price (price exclusive of taxes) for the energy and CO_2 tax, amounted to 2.23 for gasoline, which corresponds to SEK 3.61/litre. The CO_2 tax amounted to 0.43. Increasing this amount by 100 percent implies an increase of the total excise duty (or implicit tax rate) on gasoline from 2.23 to 2.66. The effect on the consumer price is an increase of 13.3 percent. Further details about the calculation of the price change can be found in [4].

Since we do not know the production function for air, bus, and train transport, we apply the same assumptions regarding energy use for these transport modes as in [4]. This means that we assume that 20 percent of the price for bus and train transport consists of energy costs (fossil fuel); the corresponding figure for air transport is 30 percent. These assumptions imply that the price for bus and train transport increases by 5.0 percent and air transport by 7.7 percent.

In Table 5 we present four different welfare measures, denoted by EV and CS_1 to CS_3. The first measure in column 1, EV, is the exact welfare measure derived in equation (3). In the second column we report the change in consumers' surplus for the trip equation, based on the parameter estimates from the restricted model. If the income effect had been zero, there would have been no difference between EV and CS_1. As the estimated income coefficient is relatively small, $\gamma = 0.012$, the difference between the values in columns 1 and 2 is also small. The measure in column 2 is given by

$$CS_1 = \frac{1}{\beta_{11}}\left[\exp(\boldsymbol{\alpha}_1 + \beta_{11}p_1^c + \gamma y + \delta_1 l) - \exp(\boldsymbol{\alpha}_1 + \beta_{11}p_1^0 + \gamma y + \delta_1 l)\right].$$

The same type measure is also presented in the third column, CS_2, but in this case we use the parameter estimates from the unrestricted model, including the effect from the cross prices. The measure in column 3 is accordingly given by

$$CS_2 = \frac{1}{\beta_{11}}[\exp(\boldsymbol{\alpha}_1 + \beta_{11}p_1^c + \beta_{12}p_2 + \gamma_1 y + \delta_1 l) - \exp(\boldsymbol{\alpha}_1 + \beta_{11}p_1^0 + \beta_{12}p_2 + \gamma_1 y + \delta_1 l)],$$

and considers only the effect of the number of trips. Finally, in the fourth column we account for the reduction in the welfare loss due to the substitution towards longer stays, according to formula (4).

As Table 5 reveals, all four welfare measures show the same pattern for the different household categories; the difference is in the level of the welfare loss. If we start the analysis by studying EV, we see that the value of this measure is slightly less than CS_1, which is expected with a small positive income effect. The difference between the two measures amounts to SEK 0.30 or 0.4 percent, evaluated at the mean of the total sample. For the income categories, the results suggest that higher income groups have a higher welfare loss than lower income groups. For households in the highest income class the welfare loss amounts to approximately SEK 80, while the figure is SEK 60 for households in the lowest income group. However, if we relate the welfare loss to the household's income, we see from the last column that the tax reform is regressive in the sense that low income households will carry a larger proportion of the tax burden in relation to household income.

For single-adult households with and without children, the difference in welfare loss is relatively small. Compared to households with two adults, the welfare loss is at about the same level as for families with three or more children. However, relating the welfare loss to income, we see that the tax burden is approximately twice as large for single-adult households with children as it is for two-adult households with children. For households with two adults and no children, the tax burden (SEK 80) is the same as for households in the highest income group. As a result of a less frequent travel behavior for families with children, the results suggest a lower welfare loss as the number of children increases in families with two adults. For travelers to the different destinations, the results indicate that travelers to Norrbotten receive the highest welfare loss, both in absolute terms and in relation to income.

Using the same type of welfare measure as in column 2, but the parameter estimates from the unrestricted model, the welfare loss is reduced by SEK 5.20 or 7.3 percent (the difference between CS_1 and CS_2). As can be seen from the table, the difference between CS_1 and CS_2 increases with income. For the lowest income group the values are equal, while the difference amounts to SEK 10.40 or 12.9 percent for the highest income group. The results also reveal that there is a smaller difference between CS_1 and CS_2 for households with one adult, compared to households with two adults.

If we consider the effects of the substitution towards longer stays, the difference between CS_2 and CS_3, the welfare loss, is reduced by an additional SEK 5.20. Thus, if we use CS_3 as a measure of the welfare loss, the average loss is reduced by 15 percent or SEK 10.10 per month compared to EV. As the table reveals, there is a relatively large difference in substitution possibilities for the different household categories. For example, the reduction in welfare loss due to longer stays amounts to only SEK 1.00 SEK for households with two adults with and without children, while it amounts to about SEK 13 for households with one adult. The results also suggest that low income households have a greater substitution possibility than high income households. For the two lowest income groups, the difference between CS_2 and CS_3 amounts to SEK $10.30 - 8.70$, while the corresponding figure is SEK $2.00 - 1.50$ for the two highest income groups. Thus the time constraints generally faced by the workforce do seem to affect households' possibilities to reduce the negative effects of increased CO_2 taxes.[5]

[Table 5 about here]

Aggregating the household-specific numbers for the last 12 months in the sample to a national level (using projected household weights), the welfare loss measured as the change in consumer surplus amounts to SEK 280 million per year when we account for the length of the visits and the substitution toward longer stays (i.e., CS_3). Compared to the change in consumer surplus

[5] As a result of the increase in the CO_2 tax, the estimated mean number of trips in the unrestricted model decreases from 1.496 to 1.486, whereas the positive cross-price effect in the day equation results in an increase of the mean number of days from 3.515 to 3.654.

from the restricted model which does not account for this substitution possibility (CS_1), CS_3 is 22 percent smaller. At an aggregate level CS_1 amounted to SEK 360 million.

6 Discussion and conclusions

In this paper we have studied the demand and welfare effects of an increased carbon dioxide tax affecting recreational travel behavior. Since a large number of countries have committed themselves to reduce their emissions of carbon dioxide in accordance with the Kyoto Agreement, or as a result of national commitments, this paper can be seen as one input in the evaluation of such a policy. In the previous literature on emission reductions the main focus has been on efficiency issues, with relatively little attention paid to distributional questions. Earlier studies accounting for distributional effects have mainly focused on aggregated effects of emission reductions based on households' total consumption of non-durable goods. In contrast the current paper analyses the effects stemming from changes in recreational travel behavior accounting for substitution possibilities neglected in the previous studies with focus on aggregated total effects.

The focus in this paper has been on recreational demand, and on the welfare and distributional effects that increased CO_2 taxes cause households. In the modeling framework we have considered households' choice of the number of trips and number of days on vacation as a simultaneous choice, where both trips and days create utility for the household. The simultaneous choices result in a non-linear count data demand system, which has been estimated using a bivariate zero-inflated Poisson lognormal model. The model is flexible and allows for both positive and negative correlation between the count data variables, in contrast to most earlier studies considering recreational demand. The inflated model choice was motivated by the large number of $(0, 0)$ observations in the empirical sample. The estimation of the parameters of the model was accomplished by the use of Gauss-Hermite quadrature.

Although the integrability conditions place strong restrictions on the cross-price parameters in the non-linear demand system, we may still find the boundary welfare effects of the environmental policy by applying the welfare measures, equivalent variation and the change in consumers' surplus, where the change in consumers' surplus, given the positive substitution effect for the number of days to stay, represents a lower bound and EV an upper bound. The results indicate that, by accounting for the number of days on vacation, the welfare loss for the households decreases by 22 percent. The exact welfare measure equivalent variation over estimates accordingly the welfare loss since it does not account for the substitutions toward longer trips.

From a distributional point of view, both measures indicate the same pattern. In the income dimension the results suggest that higher income households have a higher welfare loss measured in SEK. However, if we set the welfare loss in relation to the household's income, we see that the tax reform is regressive, in the sense that low income households carry a larger burden of the tax reform. The results also suggest that single-adult households with and without children carry a larger burden than households with two adults with children.

Appendix A: Gauss hermite quadrature

Gauss-Hermite quadrature is utilized to evaluate the integrals in this paper (equations 5 and 6). A one-dimensional integral can be obtained by factorization of $f(\varepsilon_1\varepsilon_2)$ into a conditional and a marginal distribution. Noting that $\varepsilon_1|\varepsilon_2 \sim N(\rho\varepsilon_2/\sigma_2, 1-\rho^2)$, the one-dimensional integral is given by.

$$\begin{aligned} f(x_{1h}, x_{2h}|\mathbf{z}_{1h}, \mathbf{z}_{2h}) &= \int f(x_{1h}|\exp(z'_{1h}\beta_1 + \rho\varepsilon_{2h}/\sigma_2))f(x_{2h}|\exp(z'_{2h}\beta_2 + \varepsilon_{2h}))f(\varepsilon_2)d\varepsilon_2 \\ &= \int_{-\infty}^{\infty} f(x_{1h}|\exp(z'_{1h}\beta_1 + \rho\varepsilon_{2h}/\sigma_2))f(x_{2h}|\exp(z'_{2h}\beta_2 + \varepsilon_{2h})) \\ &\quad \times \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}(\frac{\varepsilon_2}{\sigma_2})^2} d\varepsilon_2. \end{aligned}$$

The approximation with Gauss-Hermite quadrature is obtained by a change of variable. Define $\nu_h = \varepsilon_{2h}/\sigma_2\sqrt{2}$, then the equation may be written as

$$\begin{aligned} f(x_{1h}, x_{2h}|\mathbf{z}_{1h}, \mathbf{z}_{2h}) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(x_{1h}|\exp(z'_{1h}\beta_1 + \rho v_h\sqrt{2})) \\ &\quad \times f(x_{2h}|\exp(z'_{2h}\beta_2 + v_h\sigma_2\sqrt{2}))e^{(-v_h^2)}dv_h \\ &= \frac{1}{\sqrt{\pi}} \sum_{h=1}^{H} h(w_h)g(v_h) \end{aligned}$$

where $h(w_h) = f(x_{1h}|\exp(z'_{1h}\beta_1 + \rho v_h\sqrt{2}))f(x_{2h}|\exp(z'_{2h}\beta_2 + v_h\sigma_2\sqrt{2}))$ and $g(v_h) = e^{(-v_h^2)}$. Weight factors, $g(v_h)$, and abscissas, w_h, for 20-point quadrature are obtained from [1].

References

[1] Abramowitz, M. and Stegun, I.A, *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series, 55, Washington DC, 1964.

[2] Berman, M.D. and Kim, H.J. Endogenous On-site Time in the Recreation Demand Model. *Land Economics,* 75, (1999), 603-619.

[3] Browning, M. and Megihr, C. The Effects of Male and Female Labour Supply on Commodity Demands, *Econometrica,* 59, (1991), 925-951.

[4] Brännlund, R. and Nordström, J. Carbon Tax Simulations Using a Household Demand Model, *European Economic Review*, 48, (2004), 211-233.

[5] Chib, S. and Winkelmann, R. Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business and Economic Statistics,* 19**,** (2001), 428-435.

[6] Englin, J., Holmes, T. and Niell, R. Alternative Models of Recreational Off-Highway Vehicle Site Demand, *Environmental and Resource Economics*, 35(4), (2006), 327-338.

[7] Englin, J., Boxall, P. and Watson, D. Modeling Recreation Demand in a Poisson System of Equations: An Analysis of the Impact of International Exchange Rates, *American Journal of Agricultural Economics,* 80, (1998), 255-263.

[8] Englin, J. and Shonkwiler, J.S. Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand under Conditions of Endogenous Stratification and Truncation, *Review of Economics and Statistics*, 77, (1995), 104-112.

[9] Epstein, L. Disaggregate Analysis of Consumer Choice under Uncertainty, *Econometrica,* 47, (1975), 877-92.

[10] Epstein, L. Integrability of Incomplete Systems of Demand Functions, *Review of Economic Studies,* 49, (1982), 411-425.

[11] Feather, P. and Shaw, W.D. Possibilities for Including the Opportunity Cost of Time in Recreation Demand Systems. *Land Economics,* 75, (1999), 592-602.

[12] Hellström, J. A Bivariate Count Data Model for Household Tourism Demand, *Journal of Applied Econometrics,* 21, (2006), 213-226.

[13] Hellström, J. and Nordström, J. A Count Data Model with Endogenous Household Specific Censoring: The number of days to stay. *Empirical Economics*, 35, (2008), 179-192.

[14] Hicks, J.R. Consumers' Surplus and Index-Numbers, *Review of Economics Studies*, 9, (1942), 126-137.

[15] Gurmu, S. and Trivedi, P.K. Excess Zeros in Count Models for Recreational Trips. *Journal of Business and Economic Statistics*, 14, (1996), 469-477.

[16] LaFrance, J.T. and Hanemann, W.M. The Dual Structure of Incomplete Demand Systems,

American Journal of Agricultural Economics, 71, (1989), 262-274.

[17] Larson, D.M. Joint Recreation Choices and Implied Values of Time, *Land Economics,* 69, (1993), 270-286.

[18] McConnell, K.E. On-site Time in the Demand for Recreation. *American Journal of Agricultural Economics,* 74, (1992), 918-925.

[19] Morey, E., Shaw, W. and Rowe, R. A Discrete-Choice Model for Recreational Participation, Site Choice, and Activity Valuation when Complete Data are Not Available, *Journal of Environmental Economics and Management*, 20, (1991), 181-201.

[20] Munkin, M.K. and Trivedi, P.K. Simulated Maximum Likelihood Estimation of Multivatiate Mixed-Poisson Regression Models, with Application. *Econometrics Journal,* 2, (1999), 29-48.

[21] Mäler, K. Environmental Economics: A Theoretical Inquiry, Johns Hopkins University Press, Baltimore, 1974.

[22] Ozuna, T. and Gomez, I.A. Estimating a System of Recreation Demand Functions Using a Seemingly Unrelated Poisson Regression Approach, *Review of Economics and Statistics*, 76, (1994), 256-360.

[23] Shonkwiler, J.S. and Harris, T.R. Rural Retail Thresholds and Interdependencies, *Journal of Regional Science*, 36, (1996), 617-630.

[24] Train, K.E. Recreation Demand Models with Taste Differences Over People, *Land Economics*, 74, (1998), 230-239.

[25] Wang, P. A bivariate zero-inflated negative binomial regression model for count data with excess zeros. *Economics Letters,* 78, (2003), 373-378.

[26] Winkelmann, R. Health Care Reform and the Number of Doctor Visits: An Econometric Analysis. *Journal of Applied Econometrics,* 19, (2004), 455-472.

Table 1: Distribution of trips and days.

Trips	Percent	Days*	Percent
0	69.7	0	69.7
1	20.9	1	5.2
2	5.4	2	7.2
3-30	4.0	3	4.1
		4	3.9
		5	2.4
		6	1.9
		7-30	5.6

*Number of days based on the two latest trips.

Table 2: Descriptive statistics for the explanatory variables.

	Mean	Stand.dev.
Transportation cost	389.89	69.75
Cost at location	552.29	112.77
Income*	242.84	116.33
Destination dummy Stockholm	0.04	0.21
Destination dummy Gothenburg	0.03	0.18
Destination dummy Malmo	0.02	0.13
Destination dummy Norrland	0.01	0.10
Destination dummy Dalarna	0.02	0.15
Dummy for home worker	0.02	0.12
Dummy for full-time worker	0.61	0.49
Dummy for part-time worker	0.13	0.34
Dummy for students	0.19	0.39
Dummy for unemployed	0.05	0.22
Dummy for military service	0.00	0.03
Age	41.53	11.82
Number of children aged 0–6	0.21	0.53
Number of children aged 7–12	0.26	0.57
Number of children aged 13–18	0.25	0.54
Transportation mode dummy airplane	0.01	0.10
Transportation mode dummy car	0.22	0.41
Transportation mode dummy train	0.03	0.18
Transportation mode dummy bus	0.02	0.13
Number of adults in the household	1.67	0.56
Dummy for purpose: visiting relatives/friends	0.14	0.35
Dummy for purpose: visiting vacation home	0.04	0.20
Dummy for July	0.13	0.33

*Income measured in SEK thousands

Table 3: Estimation results -Restricted model.

Variable	x_1	s.e.	x_2	s.e.	Variable	π	s.e
$p_{transport}$	-0.034*	(0.016)	-	-	p_{t+l}	0.162*	(0.015)
$p_{location}$	-	-	-0.045*	(0.015)	*income*	-0.310*	(0.019)
income	0.012	(0.010)	0.012	(0.010)	*d_home worker*	-0.068	(0.137)
d_gothenburg	0.014	(0.044)	-0.001	(0.035)	*d_part-time worker*	-0.084	(0.050)
d_malmo	0.065	(0.045)	-0.067	(0.046)	*d_student*	-0.344*	(0.046)
d_norrbotten	0.108	(0.060)	0.136*	(0.063)	*d_unemployed*	-0.008	(0.078)
d_dalarna	-0.102*	(0.049)	0.140*	(0.046)	*d_military service*	0.280	(0.501)
d_home worker	-0.045	(0.127)	0.170*	(0.082)	*age*	0.149*	(0.015)
d_part-time worker	0.042	(0.039)	0.077*	(0.036)	$n\_children0-6$	0.658*	(0.325)
d_student	0.081*	(0.036)	0.222*	(0.030)	$n\_children7-12$	0.034	(0.306)
d_unemployed	0.016	(0.065)	0.099	(0.055)	$n\_children13-18$	0.892*	(0.317)
d_military service	0.199	(0.634)	-0.083	(0.543)	*Constant*	-0.554*	(0.130)
age	-0.008	(0.010)	0.028*	(0.011)			
$n\_children0-6$	-0.551	(0.312)	0.052	(0.240)			
$n\_children7-12$	-0.899*	(0.282)	0.358	(0.202)			
$n\_children13-18$	-0.535*	(0.231)	-0.187	(0.212)			
d_air	-0.321*	(0.140)	0.138*	(0.060)			
d_train	-0.238*	(0.059)	0.089*	(0.037)			
d_buss	-0.178*	(0.067)	-0.117*	(0.052)			
n_adults	-0.370	(0.253)	-0.443	(0.325)			
d_friends/family	0.091*	(0.030)	0.044	(0.025)			
d_vacation home	0.498*	(0.034)	0.401*	(0.035)			
d_july	0.060*	(0.029)	0.597*	(0.024)			
Constant	0.531*	(0.077)	1.086*	(0.074)			
σ	0.610*	(0.022)					
ρ	0.182*	(0.021)					
Log-likelihood	-34 405						

* Significant at the 5 percent level, *d_*=dummy, *n_*=number of. The restrictions are $\gamma_i = \gamma_j$ and $\beta_{ij} = \beta_{ji} = 0$.

Table 4: Estimation results unrestricted model.

Variable	x_1	s.e.	x_2	s.e.	Variable	π	s.e
$p_{transport}$	-0.016	(0.016)	0.083*	(0.016)	p_{t+l}	0.155*	(0.015)
$p_{location}$	-0.106*	(0.020)	-0.076*	(0.016)	*income*	-0.306*	(0.019)
income	0.042*	(0.015)	0.027*	(0.013)	*d_home worker*	-0.018	(0.137)
d_gothenburg	0.016	(0.044)	0.013	(0.035)	*d_part-time worker*	-0.077	(0.050)
d_malmo	0.064	(0.046)	-0.068	(0.046)	*d_student*	-0.345*	(0.046)
d_norrbotten	0.091	(0.060)	0.057	(0.064)	*d_unemployed*	-0.008	(0.078)
d_dalarna	-0.074	(0.050)	0.164*	(0.046)	*d_military service*	0.237	(0.493)
d_home worker	-0.035	(0.129)	0.231*	(0.081)	*age*	0.144*	(0.015)
d_part-time worker	0.026	(0.040)	0.077*	(0.036)	$n\_children0-6$	0.497	(0.325)
d_student	0.068	(0.037)	0.225*	(0.030)	$n\_children7-12$	0.072	(0.307)
d_unemployed	0.032	(0.066)	0.092	(0.056)	$n\_children13-18$	0.872*	(0.317)
d_military service	0.177	(0.566)	-0.068	(0.495)	*Constant*	-0.480*	(0.130)
age	-0.020	(0.011)	0.022	(0.011)			
$n\_children0-6$	-0.304	(0.315)	0.042	(0.241)			
$n\_children7-12$	-0.718*	(0.288)	0.365	(0.202)			
$n\_children13-18$	-0.472*	(0.233)	-0.227	(0.212)			
d_air	-0.348*	(0.140)	0.089	(0.061)			
d_train	-0.230*	(0.059)	0.081*	(0.037)			
d_buss	-0.193*	(0.069)	-0.132*	(0.052)			
n_adults	0.575	(0.362)	-0.502	(0.326)			
d_friends/family	0.085*	(0.031)	0.046	(0.025)			
d_vacation home	0.426*	(0.034)	0.368*	(0.035)			
d_july	0.054	(0.029)	0.606*	(0.024)			
Constant	0.855*	(0.100)	0.930*	(0.095)			
σ	0.609*	(0.022)					
ρ	0.176*	(0.022)					
Log-likelihood	-34 372						

* Significant at the 5 percent level, *d_*=dummy, *n_*=number of

Table 5: Mean welfare effect for different household categories

		EV	CS_1	CS_2	CS_3	EV/Inc[a]
Income	0-150	59.6	59.9	59.1	48.8	0.77
in SEK thousand	151-210	67.3	67.6	64.8	56.1	0.38
	211-280	71.5	71.8	66.8	63.2	0.29
	281-350	74.6	75.0	68.0	66.0	0.24
	351-785	80.4	80.8	70.4	68.9	0.19
One-adult households	-without children	63.4	63.6	61.8	49.9	0.57
	-with children	63.9	64.2	62.5	47.8	0.47
Two-adult households	-without children	80.4	80.9	73.1	71.9	0.31
	- 1 child	74.8	75.2	68.0	66.8	0.27
	- 2 children	68.3	68.6	62.3	61.1	0.23
	- 3 or more children	64.6	64.8	60.3	59.0	0.24
Destination	Stockholm	63.9	64.2	59.6	54.4	0.34
	Gothenburg	65.2	65.5	60.1	55.0	0.32
	Dalarna	69.8	70.1	64.6	60.1	0.34
	Malmo	65.2	65.5	60.1	55.3	0.38
	Norrbotten	94.3	94.9	90.8	82.7	0.54
Mean for total sample		70.7	71.0	65.8	60.6	0.37

EV equivalent variation. CS_1 consumers' surplus integrability restricted trip demand equation. CS_2 consumers' surplus unrestricted trip demand equation. CS_3 consumers' surplus unrestricted demand system (trip and day equations). [a] inc = income in thousand SEK.