Using a Control Function to Resolve the Travel Cost Endogeneity Problem in Recreation Demand Models

Melstrom, Richard and Lupi, Frank

Salisbury University, Michigan State University

August 2012

Using a Control Function to Resolve the Travel Cost Endogeneity Problem in Recreation Demand Models

Richard T. Melstrom
Department of Economics and Finance
Department of Environmental Studies
Salisbury University

Frank Lupi
Department of Agricultural, Food and Resource Economics
Department of Fisheries and Wildlife
Michigan State University

Running title: Resolving Travel Cost Endogeneity with a Control Function

Corresponding Author: Richard T. Melstrom
Department of Economics and Finance
Salisbury University
Purdue Hall
1101 Camden Avenue
Salisbury, MD 21801-6860
rtmelstrom@salisbury.edu

Using a Control Function to Resolve the Travel Cost Endogeneity Problem in Recreation Demand Models

Abstract

This paper proposes using a control function to correct for endogeneity in recreation demand models. The control function approach is contrasted with the method of alternative specific constants (ASCs), which has been cautiously promoted in the literature. As an application, we consider the case of travel cost endogeneity in the demand for Great Lakes recreational fishing. Using data on Michigan anglers, we employ a random utility model of site choice. We show that either ASCs or the control function can correct for travel cost endogeneity, although we find that the model with ASCs produces significantly weaker results. Overall, compared with traditional approaches control functions may offer a more flexible means to eliminate endogeneity in recreation demand models.

Keywords

Recreation demand, random utility model, travel cost method, travel cost endogeneity, control function, alternative specific constants, recreational fishing

1 Introduction

How the demand for recreation trips is formulated can play a significant role in valuing recreation sites and site quality characteristics. In particular, if some trip and site characteristics important to the recreationist's visiting decision are ignored (e.g. water quality, substitute sites), researchers may misinterpret the value of other, observed site characteristics or the value of the site itself. How a researcher chooses to addresses the site characteristics not directly observed is therefore an important concern.

Endogeneity of explanatory variables is a standard econometric problem. In recreation demand modeling, endogeneity is typically a matter of omitted variables, in that the observed, explanatory variables may be correlated with the unobservables. For example, hikers may be willing to drive farther in order to hike remote trails, inducing a positive correlation between remoteness and travel cost, so if remoteness is unobserved travel cost will be endogenous. Endogeneity violates a critical consistency condition in model estimation, so recreation demand models that ignore omitted variables risk biased parameter and welfare estimates (Moeltner and von Haefen 2011). This problem is generally recognized by recreation demand researchers (e.g. Parsons 1991; Hausman et al. 1995; Murdock 2006; Weber et al. 2012). Biased estimates of the travel cost effect in recreation demand is also a particularly serious issue because the travel cost coefficient is the denominator in formulas for the valuation of sites and site characteristics.

This paper explores the issue of travel cost endogeneity in a model of the demand for recreational fishing. Using a random utility framework and data on site characteristics and angler trips, one can value recreational fishing sites and their amenities, but these values will be biased if travel cost is endogenous. We examine and compare two methods to correct for travel cost endogeneity. The first method is alternative specific constants (ASCs), which has been used in recreation demand applications before (e.g. Murdock 2006; Hynes et al. 2007). The second method is the control function, which is novel to recreation demand modeling. A control function uses excluded instruments in a two-step estimation approach akin to two-stage least squares (2SLS). Using time variation in fuel prices, we derive an instrument that effectively conditions out the role omitted variables play on travel cost.

Our application uses data on Great Lakes recreational anglers. Catch rates unique to the timing of fishing trips are used to describe the quality of fishing sites. Catch rates therefore vary across alternatives and time, so the fixed-effects nature of ASCs does not prohibit estimation of the catch rate effects in the site choice model. We examine the significance of travel cost endogeneity by comparing parameter and welfare estimates from models which ignore correlation between travel cost and the unobservables with those that use ASCs or a control function to correct for endogeneity. We also compare the usefulness of ASCs and control functions by comparing the results from the models which test these methods. Overall, our results indicate that applied researchers may find the control function a flexible means of correcting for endogeneity in recreation demand models.

2 Background

The travel cost endogeneity issue was first raised by Parsons (1991) and Randall (1994), who noted that recreation preferences could influence the choice of residential location. That is, travel cost is not exogenous if a person considers the distance to the recreation site when they choose where to live. This argument is closely related to the theory motivating hedonic analysis, which values amenities through home prices. With hedonic analysis, individuals are assumed to select a home based on the proximity of work and leisure opportunities and environmental amenities. In so far as commuting costs to activities—including recreation—are a significant determinant of residential location, models that assume travel cost is exogenous are inconsistent with household behavior.[1] In general, though, there are any number of reasons an unobserved recreational site amenity could be correlated with travel cost and lead to an endogeneity problem.

Some evidence does, in fact, suggest that travel cost could be correlated with site amenities. Phaneuf et al. (2008) study home buyers in North Carolina and find that access to recreation sites and water quality at recreation sites are important factors in residential location. Albouy (2009), in a study of household preferences for metropolitan areas of the United States, finds that households are willing to pay to live near the coasts, including the Great Lakes. Deller et al. (2001) look at the characteristics of rural areas in the United States and find that a county's recreational infrastructure, water and land features are major determinants of rural population growth. With specific regard to anglers, Weithman and Haas (1982) report over one-fourth of anglers in Taney County, Missouri chose their home partly due to nearby fishing opportunities.

The handful of recreation demand studies investigating endogeneity present an assortment of the strategies that could be used to correct for the problem in different demand modeling contexts. Parsons (1991) uses a 2SLS method in a single site application, and finds that ignoring travel cost endogeneity biases estimates of site value downward. Despite this significant finding, the matter of travel cost endogeneity received little attention until recently. This may be due to the limitations of 2SLS, which cannot be used in the nonlinear site choice models that most recreation demand modelers now use (Moeltner and von Haefen 2011).

[1] Sometimes this issue can be ignored without consequence in recreation demand modeling. Specifically, travel cost may be endogenous in the strict sense but not lead to a problem econometrically if travel cost is only correlated with the observed site characteristics and not the unobserved site characteristics. Also, if the amenities that affect residential location are thought to be largely uncorrelated with the site characteristics important to the recreation decision, then travel cost endogeneity may be less of a concern.

One approach to dealing with travel cost endogeneity in site choice models is the inclusion of ASCs, which are site-specific fixed effects. ASCs absorb all site-specific effects, both observed and unobserved, to eliminate the omitted variables problem. Hausman et al. (1995) use fixed effects in this manner in a model of recreational trips in Alaska. Generally, though, ASCs are rarely employed in site choice models, perhaps because of the limitations they impose in cross-section samples (Englin and Cameron 1996). That is, in site choice modeling ASCs preclude identification of the site-specific effects—a problem avoided by Hausman et al. who had panel data on site visits and site characteristics, such as catch rates, that changed over time.

Murdock (2006) offers a solution to this limitation by demonstrating that, after estimating a site choice model with ASCs, the coefficients of the site-specific characteristics can be recovered in an auxiliary regression. In essence, this is possible because the estimated ASCs preserve the identifying variation of the site-specific effects. The framework for this method was originally developed by Berry (1994) and Berry et al. (1995) for models of differentiated product markets. Murdock suggests the following two stage approach. First, estimate the choice model with ASCs. If travel cost is correlated with unobserved site-specific characteristics, the effect of travel cost in the choice model is now estimated consistently. Second, regress the estimated ASCs on the observed site-specific characteristics (using 2SLS if appropriate). The site-specific effects are thus estimated in the second stage. The development of this bias-correction method has encouraged the recent use of ASCs in recreation demand models (e.g. Phaneuf et al. 2009; Timar and Phaneuf 2009; Jakus et al. 2010).

Von Haefen and Phaneuf (2008) and Moeltner and von Haefen (2011) caution against the use of the two stage ASCs method to correct for endogenous site characteristics. They argue that the effectiveness of the approach relies on features of the data and model, such as the use of a large choice set with a sufficient number of visits to each alternative. However, von Haefen and Phaneuf (2008) show that the use of stated preference data can alleviate some of the drawbacks of the method.[2]

Alternatively, a control function can be used to correct for endogeneity. A control function works by allowing a model to condition on the part of the unobservables correlated with the observed covariates (Petrin and Train 2006). Employing a control function is similar to standard 2SLS methods, although it remains valid in nonlinear models. Early applications of a control function were performed by Smith and Blundell (1986) in a tobit model and Rivers and Voung (1988) in a probit model. More recent applications include Liu et al. (2011) and Ricker-Gilbert et al. (2011). Petrin and Train (2010) present the control function approach as an alternative to the bias-correction methods of Berry (1994) and Berry et al. (1995), thus it may be a useful alternative to the two stage ASCs method in recreational site choice models. The approach also can be used in situations where the ASCs method is unsound. As with 2SLS, however, a good instrumental variable is necessary for the control function to eliminate endogeneity (Petrin and Train 2010).

3 Model

3.1 Site Choice Model

We develop a random utility model in the context of a recreational angling site choice problem. Consider an angler's utility function. On a fishing trip occasion, an angler maximizes utility by selecting the site which delivers the greatest utility subject to an income constraint. Let y_i be the

[2] Recent work by Abidoye et al. (2012) examine a Bayesian approach to recover the effects of site-specific characteristics when ASCs are used to control for unobserved site characteristics. Unlike the two stage estimation approach of Murdock (2006), the approach suggested by Abidoye et al. can be used in general applications of mixed logit models.

income of angler i, p_{ikt} the travel cost to site k at time t, $\boldsymbol{q_k}$ a vector of quality measures that vary across sites and $\boldsymbol{q_{kt}}$ a vector of quality measures that vary across sites and time. The vector $\boldsymbol{q_k}$ includes site-specific characteristics, such as a variable measuring the shoreline length of a fishing site. The vector $\boldsymbol{q_{kt}}$ includes site characteristics that change with time, such as fish catch rates. Assume utility is linear in its arguments and has a marginal utility of money that is constant across the choice alternatives. Then, after substituting in the budget constraint, angler i's utility conditional on visiting site k at time t is

(1) $$U_{ikt} = \rho(y_i - p_{ikt}) + \boldsymbol{\beta_1 q_k} + \boldsymbol{\beta_2 q_{kt}} + \varepsilon_{ikt}.$$

where ρ is the marginal utility of money, $\boldsymbol{\beta_1}$ is the vector of parameters on site-specific characteristics, $\boldsymbol{\beta_2}$ is the vector of parameters on site and time varying characteristics and ε_{ikt} is the unobserved utility component. The error ε_{ikt} is usually assumed to be independent of all other variables in the right-hand side of (1), but this may not be the case. In particular, travel cost endogeneity arises if p_{ikt} is correlated with ε_{ikt}, which is to say $cov(p_{ikt},\varepsilon_{ikt}) \neq 0$.[3]

3.2 Alternative Specific Constants

The ASCs method uses a vector of $K - 1$ site-specific fixed effects, with K the number of choice alternatives. Denoting the ASCs vector as $\boldsymbol{\alpha_k}$, then angler utility is transformed into

(2) $$U_{ikt} = \rho(y_i - p_{ik}) + \boldsymbol{\alpha_k} + \boldsymbol{\beta_2 q_{kt}} + \zeta_{ik},$$

where $\boldsymbol{\alpha_k}$ absorbs the observed effects $\boldsymbol{\beta_1 q_k}$ and unobserved site effects in ε_{ikt}. In this way, ASCs prevent estimating the coefficients $\boldsymbol{\beta_1}$ in a random utility model. However, assuming that the

[3] Note that ε_{ikt} could also be correlated with $\boldsymbol{q_k}$ or even $\boldsymbol{q_{kt}}$. With suitable instruments the methods examined here can apply to alternative endogeneity problems.

idiosyncratic portion of the error is uncorrelated with angler travel cost, i.e. $cov(p_{ikt},\zeta_{ikt}) = 0$, the remaining parameters in (2) can be estimated consistently.

The coefficients $\boldsymbol{\beta_1}$ may not be estimable in (2), but they can still be recovered. This is achieved by regressing the estimated ASCs ($\boldsymbol{\alpha_k}$) on the site-specific characteristics ($\boldsymbol{q_k}$). This auxiliary regression can be estimated with ordinary least squares or, if $\boldsymbol{q_k}$ is endogenous, with 2SLS (Murdock 2006).

The ASCs method does have some limitations. For sites with few visits $\boldsymbol{\alpha_k}$ is unlikely to be well identified in the site choice model, hindering identification of $\boldsymbol{\beta_1}$ in the second regression. In fact, it is not possible to estimate the constants for sites with no visits, so it is necessary to drop unvisited sites when using ASCs. Identification of $\boldsymbol{\beta_1}$ in the second regression is also difficult if there are relatively few sites in the choice set. Furthermore, poor estimation of the constants $\boldsymbol{\alpha_i}$ for sites with few visits may affect estimates of $\boldsymbol{\beta_2}$ in the first step. It may also be inappropriate to perform the two step ASCs method in mixed logit models (Abidoye et al. 2012).

3.3 Control Function

Employing a control function is similar to standard instrumental variables methods involving a two-step estimation procedure, such as 2SLS. Petrin and Train (2006) (see also Imbens and Wooldridge 2007) show that under basic assumptions one can include in a discrete choice model with endogenous covariates an additional variable to condition for the portion of the error correlated with the observables.

For the recreational fishing site choice model described above, suppose travel cost can be expressed in reduced form as

(3) $p_{ikt} = \boldsymbol{\gamma_q q_{kt}} + \boldsymbol{\gamma_z z_{ikt}} + \mu_{ikt}$,

where z_{ikt} is a vector of observed covariates and μ_{ikt} is unobserved. The argument here is that p_{ikt} is endogenous because it is in part determined by the choices of anglers. In this case, z_{ikt} are observables important to an angler's travel cost but not an angler's site choice decision (i.e. utility at a fishing site), while μ_{ikt} contains elements in the angler's travel cost unobserved by the researcher.

Assume μ_{ikt} is independent of $\boldsymbol{q_{kt}}$ and $\boldsymbol{z_{ikt}}$. Travel cost endogeneity arises because the unobserved component of travel cost, μ_{ikt}, is correlated with the unobserved component of the site utility function, ε_{ikt}. For example, this occurs because a site characteristic like scenic beauty matters in fishing site choice and residential location, but is unobserved by the researcher, so its effect becomes captured by both μ_{ikt} and ε_{ikt}. However, an estimate of μ_{ikt} can be had by estimating the travel cost equation (3) and calculating the residuals, $\hat{\mu}_{ikt}$. Inserting $\hat{\mu}_{ikt}$, which is referred to as the control function, into the random utility model,

(4) $U_{ikt} = \rho(y_i - p_{ik}) + \boldsymbol{\beta_1 q_k} + \boldsymbol{\beta_2 q_{kt}} + \varphi\hat{\mu}_{ikt} + \omega_{ikt}$.

Assuming $\hat{\mu}_{ikt}$ is a good estimate of μ_{ikt}, the control function captures the effect within ε_{ikt} that p_{ikt} is correlated with, $cov(p_{ikt},\omega_{ikt}) = 0$, and the parameters in (4) can be estimated consistently (Petrin and Train 2010). Evidence for endogeneity is confirmed if the coefficient on the control function, φ, is significantly different from zero.

In short, the control function procedure consists of two steps. First, the endogenous variable in the site choice model is regressed against the observables and a vector of instruments.

Second, the site choice model is estimated but with the residual from the first regression as an additional variable.

As with ASCs, the control function method has some limitations. First, a control function requires an instrument. These instruments, $\boldsymbol{z}_{\boldsymbol{ikt}}$, must be correlated with p_{ikt}, hence $\boldsymbol{\gamma}_z > 0$, but uncorrelated with ε_{ikt}, $cov(\boldsymbol{z}_{\boldsymbol{ikt}},\varepsilon_{ikt}) = 0$. Second, a control function for p_{ikt} will only capture endogeneity correlated with p_{ikt}. Thus, if $\boldsymbol{q}_{\boldsymbol{k}}$ is endogenous it will be necessary to include additional controls. Finally, if the estimate of φ is statistically significant, the standard errors in the site choice model stage need to be corrected for two-stage estimation.[4]

3.4 Estimation and Welfare Analysis

In (1), (2) and (4), assuming the error is independently and identically distributed extreme value yields the conditional logit (CL) model. The probability of angler i choosing alternative k at time t is

$$(5)\qquad \pi_{ikt} = \frac{e^{V_{ikt}}}{\sum_{l=1}^{K} e^{V_{ilt}}},$$

where V_{ikt} is the deterministic portion of the utility function and K is the number of choice alternatives. For the traditional CL $V_{ikt} = \rho p_{ikt} + \boldsymbol{\beta}_{\boldsymbol{1}}\boldsymbol{q}_{\boldsymbol{k}} + \boldsymbol{\beta}_{\boldsymbol{2}}\boldsymbol{q}_{\boldsymbol{kt}}$, for the CL with ASCs $V_{ikt} = \rho p_{ikt} + \boldsymbol{\alpha}_{\boldsymbol{k}} + \boldsymbol{\beta}_{\boldsymbol{2}}\boldsymbol{q}_{\boldsymbol{kt}}$ and for the CL with control function $V_{ikt} = \rho p_{ikt} + \boldsymbol{\beta}_{\boldsymbol{1}}\boldsymbol{q}_{\boldsymbol{k}} + \boldsymbol{\beta}_{\boldsymbol{2}}\boldsymbol{q}_{\boldsymbol{kt}} + \hat{\mu}_{ikt}$. The parameters of these site choice models are estimated via maximum likelihood.[5]

[4] This could be done using the bootstrap.

[5] Specifications other than conditional logit were tested, including nested logit and mixed logit. The improvement in the fit of these two models over the conditional logit was extremely modest. Moreover, these various models have been found to deliver similar welfare estimates (von Haefen 2003).

In the second regression of the ASCs method, ordinary least squares is used to estimate the site-specific effects. In the first stage of the control function approach ordinary least squares is used to estimate (3).

For welfare measurement, from the observed indirect utility function V_{ikt} the expected per-trip compensating variation measure is

(6) $$E[CV] = \frac{1}{\rho}\left[ln\left(\sum_{l=1}^{K} e^{V_k^1}\right) - ln\left(\sum_{l=1}^{K} e^{V_k^0}\right)\right]$$

where V_k^0 and V_k^1 are the conditional utility levels before and after, respectively, a change in price or quality. It is clear that a poor estimate of ρ will produce invalid site values. If the estimate of ρ is biased upwards (downwards), *ceteris paribus*, welfare estimates will be too small (large).

4 Data

4.1 Fishing site characteristics

Data are obtained from a 2008-2009 survey of anglers fishing in Michigan. Information from the survey includes the location and date of anglers' most recent trips, targeted fish species, whether the purpose of the trip was primarily to fish and how many days were spent fishing.[6]

The sample we use in the application includes only sportfishing day trips to the Great Lakes taken by Michigan residents traveling less than 200 miles one-way to their destination, which is considered the maximum feasible distance for a day trip. Accordingly, for each trip the set of choice alternatives includes only sites within 200 miles of an angler's home. Information on site quality measures is gathered independently of the survey.

[6] See Simoes (2009) for a detailed treatment of the survey and the sample.

Travel costs are calculated from distance information, angler characteristics and gasoline prices. Per-mile fuel costs are Michigan monthly retail per-gallon gasoline prices (EIA 2012) divided by 23.1, which is the typical fuel economy for a car (AAA 2009). Fuel costs are raised 33% for anglers who report trailering a boat on their trip. Per-mile maintenance and depreciation costs are collected from AAA reports (AAA 2008; AAA 2009). The opportunity cost of travel time is computed as a 1/3 an angler's reported income divided by 2000, normalized for time by assuming an average driving speed of 45 miles per hour. Travel cost is the sum of per mile fuel cost, per mile maintenance and depreciation cost and the opportunity cost of travel time.

Michigan Department of Natural Resources (MDNR) creel survey data are used to compute expected monthly per-hour catch rates at fishing locations on the Great Lakes for six fish species: chinook, coho, steelhead, lake trout, walleye and yellow perch. These are the most popular species among Great Lakes anglers. For the warm water species, the MDNR defines the catch rate as

$$catch\ rate_i = \frac{number\ of\ species\ i\ landed}{hours\ spent\ fishing\ for\ i}$$

for *i* = walleye and yellow perch. For the salmonids, the MDNR computes catch rates over an effort level that includes the time spent fishing for other salmon species.[7] That is,

$$catch\ rate_i = \frac{number\ of\ species\ i\ landed}{\sum_j hours\ spent\ fishing\ for\ j}$$

for *i, j* = chinook, coho, steelhead and lake trout. The MDNR creel data come from monthly surveys of 66 sites over several years. Due to missing average catch rates for several months at many sites, a series of tobit regressions are used to predict average catch rates, which are used as

[7] Salmon catch rates are reported in this manner because it is relatively easy for anglers to target more than one salmon species at a time .

the catch rate variables for the site choice model.[8] Table 1 provides descriptive statistics from the variables constructed from the angler and creel survey datasets.

The reported destination of each trip is matched with the closest creeled fishing site. Trips in the sample are spread over approximately two years, and the catch rates at choice alternatives vary with the month of a trip, so catch rates varied across sites and time.

In the site choice model, trips that target salmonids—chinook, coho, steelhead and lake trout—are distinguished from those targeting warm water species. This distinction is desirable because these categories, referred to as product lines, constitute different types of fishing experiences (Kikuchi 1986; Jones and Lupi 2000). Catching salmonids in the Great Lakes involves fishing in certain temperature zones, usually from boats in deep water during the summer or from breakwaters and river mouths in the spring and fall. Perch and walleye can be caught in warmer temperature zones, often in the same locations year-round. We therefore assign two product line alternatives to each fishing site, so the site choice model includes a total of 132 choice alternatives. Trips are assigned to a product line at preferred sites using the targeting information reported by anglers.

A number of dummy variables are included as covariates. The variable *Urban* takes the value one if the U.S. Census 2000 defines a portion of the choice alternative's zip code area urban. *Urban* is intended to measure services at a site. *Highway* takes the value one if a state highway runs adjacent to the alternative and is intended to measure the remoteness of a site. *Bayorseaway* takes the value one if the alternative is located on a large bay and is intended to

[8] Using an estimate of catch rates is necessary but introduces measurement error (Morey and Waldman 1998). However, the approach remains consistent and the error variance will decline as more data become available (Train et al. 2000). Since the MDNR use stratified random sampling, month-to-month, to creel fishing sites, and surveying at creeled sites is extensive, the bias should be limited. On the other hand, it is also possible that the catch rate regressions make predictions that more accurately reflect angler perceptions of catch rates than would using true catch rates. Alternative methods of filling in missing data, including multiple imputation, were found impracticable.

capture the preference for fishing in areas of the lake that are typically warmer. *Bayorseaway* is also interacted with a dummy variable indicating a warm water alternative (*Warmwater*), in order to measure how *Bayorseaway* can affect preferences for sites across the two product lines differently.

The indirect utility function for the traditional CL specification of the Great Lakes angling demand model is therefore

$$V_{ikt} = \rho p_{ikt} + \boldsymbol{\beta}_{CR}\boldsymbol{CR}_{kt} + \beta_U Urban_k + \beta_H Highway_k + \beta_B Bayorseaway_k$$
$$+\beta_{BWW} Bayorseaway_k \cdot Warmwater_k,$$

where $\boldsymbol{CR}_{kt}$ is the vector of catch rates. The indirect utility function for the CL with the control function is identical, except for the inclusion of the control function as a covariate. The indirect utility function for the CL with ASCs is $V_{ikt} = \rho p_{ikt} + \boldsymbol{\alpha_k} + \boldsymbol{\beta_{CR}CR_{kt}}$; the site-specific effects are recovered by estimating $a_k = \beta_U Urban_k + \beta_H Highway_k + \beta_B Bayorseaway_k +$ $\beta_{BWW} Bayorseaway_k \cdot Warmwater_k$.

4.2 Instrumental Variables

A control function can only correct for travel cost endogeneity with a valid instrumental variable. This instrument must be correlated with travel cost but uncorrelated with the unobservables affecting site choice. Furthermore, because the site choice model is specified as a CL, the instrument must have cross-site variation, so that the control function is not differenced away in formulating the likelihood function and estimating the model.

The nature of the instrument used for the control function is based on the construction of the travel cost variable described in § 4.1. Recall that travel cost depends on, in addition to the usual factors, monthly fuel prices. The data set is not a panel, but the sampling of trips across 2008-2009 captures several large exogenous changes in fuel prices that influence travel costs.

As an instrument, we use crude oil prices interacted with the mileage distance to fishing sites. This instrument clearly satisfies the condition that it be correlated with travel cost. It is not possible to test the exogeneity restriction. However, it is unlikely that crude oil price times distance is correlated with the unobserved site-specific effects of the site choice model, for two reasons. First, unobserved characteristics, such as scenic beauty, are not likely to be a function of or determined by (current) oil prices and distances. Second, unobserved natural events in Michigan or the Great Lakes that may influence site choice are unlikely to impact and therefore be correlated with oil prices, since oil prices are determined on a global scale. We also include month-year fixed effects in the travel cost equation to control for exogenous shifts in demand.

5 Results

We estimate four different site choice models. The first model is estimated as a traditional CL, which ignores any potential travel cost endogeneity. The second model includes only the set of alternatives that receive visits in the sample (there are 31 unvisited alternatives). The third model includes ASCs, which also does not include any unvisited sites in the choice set. In this site choice mode, the ASCs preclude estimating the site-specific effects, so the parameters of site-specific characteristics are estimated in an auxiliary regression. The final estimated model is identical to the first CL except that it includes the control function. The results are presented in table 2.

In the estimated traditional CL, the catch rate parameters are all positive, implying that trips are more frequently taken to sites with high catch rates. The parameters on *Urban* and

Highway are also positive, suggesting that the preferred sites have urban amenities and are not isolated. In contrast, the effect of *Bayorseaway* is negative while the effect of *Bayorseaway·Warmwater* is positive, implying that fishing sites in bays or seaways tend to be avoided unless the trip is taken with the intention of fishing for warm water species. This effect may arise for two reasons. First, bays and seaways offer more protected waters for anglers using smaller boats, and these tend to be warm water anglers. Second, bays or seaways provide good habitat for warm water species, and there are other warm water species anglers target but we lacked the catch rate data to include in the model. Thus, anglers that prefer warm water alternatives are drawn to these locations; the opposite holds true for anglers that prefer cold water species. The results of the CL that includes only visited alternatives are similar, although the precision of estimates is less, perhaps due to reduced variation in the site characteristics.

The CL with ASCs yields substantially weaker results compared with the two prior models. Although the catch rate parameters can be estimated in the presence of ASCs, the precision of the estimates is poor. Only the effect of the chinook salmon and walleye catch rates remain positive and significant. On the other hand, there is a large improvement in the fit of the model—the log likelihood is considerably smaller compared with the other CLs.

Table 3 presents the results of regressing the estimated ASCs on the site-specific characteristics (which in this case are all dummy variables). Compared with the other site choice models, only the effect of *Bayorseaway·Warmwater* retains the same sign and significance level. In modeling the demand for Great Lakes recreational fishing, these results, combined with the weak identification in the first step, reduce the appeal of using ASCs to control for travel cost endogeneity.[9]

[9] Although the results are not reported here, many of the estimated ASCs are not precisely estimated. This may explain the lack of identification in the second step.

Now consider the control function approach. The travel cost equation estimates are presented in table 4. The results indicate a high level of explanatory power in the instrument. The statistical significance of the month-year dummy variables also indicates the role of other shocks (shifts in regional fuel demand) during this time to travel costs. An R-squared of 0.709 indicates that the explanatory power of the equation is large. Thus, the control function should be a good estimate of the part of travel cost correlated with the unobservables in the site choice model.

The results of the CL with the control function are presented in the final column of table 2. The results are largely similar to the results of the traditional CL, except for the new covariate. The t-statistic on the control function coefficient indicates statistically significant travel cost endogeneity. Note that the travel cost coefficient in the model with ASCs is nearly equivalent to its counterpart in the model with the control function, although the change in the travel cost coefficient in these two models compared with the traditional CL is quite small – less than 5%.

Correcting for travel cost endogeneity increases, in absolute value, the travel cost coefficient. This bias would lead to welfare estimates that are too high, which is confirmed by comparing the welfare impacts of some plausible scenarios, reported in tables 5 and 6. In table 5, the welfare measures reflect the average per Great Lakes trip willingness to pay (WTP) to forgo the suggested change in catch rates across all sites. Table 6 reports the average per trip WTP among the trips taken to a particular Great Lake to forgo loss of access to that lake. In every scenario, the traditional CL model predicts a greater welfare impact compared with the model with the control function. However, the level of bias in welfare estimates is relatively small.

In fact, whether the model with ASCs or the control function is preferred will have a greater impact on welfare predictions than the issue of travel cost endogeneity per se. For example, the chinook catch rate coefficient is 7.996 is the model with ASCs but 12.018 in the model with the control function. As a result, the per trip WTP to forgo a 50% decrease in chinook catch rates is lower in the model with ASCs ($2.14 vs. $3.32). Possibly, some of the catch rate coefficients in the model with ASCs differ from their counterparts in the other models because the ASCs also control for other omitted variables. The loss of identifying power in the model with ASCs—because useful cross-site variation in catch rates is absorbed by the ASCs—makes it difficult to determine whether catch rates are truly endogenous.

6 Conclusion

The travel cost method is a useful means of estimating the value of recreation sites and site amenities, but controlling for endogeneity in recreation demand models, particularly random utility models, can be difficult. Several solutions to the endogeniety problem can be found in the recreation demand literature. Unfortunately, all of the suggested corrections rely on stringent model and data requirements, or require a complicated estimation procedure. In contrast, a control function, which conditions out the endogeneity problem in nonlinear models using an intuitive two-step estimation procedure akin to 2SLS, is much less restrictive.

We use a control function to correct for travel cost endogeneity bias in a site choice model of Great Lakes recreational fishing. The demand model uses two sources of data that have a useful time element: a sample of trips spread over two years and choice alternatives with catch rates that vary by month. To derive the control function, we use monthly oil prices to instrument for travel cost. Estimating the site choice model with the control function reveals statistically significant evidence that travel cost is endogenous.

The results from the model with the control function are compared with a site choice model that employs ASCs to correct for endogeneity. Catch rates vary across sites and trips, so the catch rate parameters can be estimated even with ASCs in the site choice model. However, we find running the model with ASCs yields imprecisely estimated catch rate effects. Furthermore, the site-specific effects are not well identified when the ASCs are regressed on the site-specific characteristics. In contrast, we find that the control function corrects for travel cost endogeneity and preserves identification of all the parameters.

We find ignoring travel cost endogeneity biases the travel cost parameter estimate downward toward zero, a result which is robust because both the model with ASCs and the model with the control function identify the same level of bias. However, compared with the point estimate of the parameter, the bias is small, and ignoring the endogeneity does not meaningfully skew estimates of WTP. These results suggest that travel cost endogeneity may not be a problem in valuing Great Lakes recreational fishing, although this conclusion may not extend to other recreational activities. Von Haefen and Phaneuf (2008), Jakus et al. (2010) and Abidoye et al. (2012), in models of moose hunting, off-highway driving and lake recreation, respectively, all find that the travel cost coefficient is biased toward zero and that the magnitude of the bias is large. Thus, testing and correcting for travel cost endogeneity will continue to be important in recreation demand modeling.

7 References

Abidoye, B.O., J.A. Herriges and J.L. Tobias (2012). Controlling for Observed and Unobserved Site Characteristics in RUM Models of Recreation Demand. *American Journal of Agricultural Economics* 94: 1070-1093.

Albouy, D. (2009). What are Cities Worth? Land Rents, Local Productivity, and the Value of Amenities. NBER Working Paper No. 14981.

American Automobile Association (AAA). *Your Driving Costs: 2008 Edition*.

American Automobile Association (AAA). *Your Driving Costs: 2009 Edition*.

Berry, S.T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics* 25: 242-262.

Berry, S., J. Levinson and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica* 63: 841-890.

Deller, S.C., S.T. Tsung-Hsiu, D.W. Marcouiller and D.B.K. English (2001). The Role of Amenities and Quality of Life In Rural Economic Growth. *American Journal of Agricultural Economics* 83: 352-365.

Energy Information Administration (EIA). *Michigan Total Gasoline Through Company Outlets Price by All Sellers (Dollars per Gallon)*. Accessed online at http://tonto.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMA_EPM0_PTC_SMI_DPG&f=M.

Englin, J. and T.A. Cameron (1996). Augmenting Travel Cost Models with Contingent Behavior Data. *Environmental and Resource Economics* 7: 133-147.

Hausman, J.A., G.K. Leonard and D. McFadden (1995). A Utility-Consistent, Combined Discrete Choice and County Data Model Assessing Recreational Use Losses Due to Natural Resource Damage. *Journal of Public Economics* 56: 1-30.

Hynes, S., N. Hanley and E. Garvey (2007). Up the Proverbial Creek without a Paddle: Accounting for Variable Participant Skill Levels in Recreational Demand Modelling. *Environmental and Resource Economics* 36: 413-426.

Imbens, G.W. and Wooldridge, J.M. (2007). What's New in Econometrics. Unpublished notes, National Bureau of Economic Research, Summer Institute 2007. Accessed online at http://www.nber.org/WNE/lect_1_match_fig.pdf.

Jakus, P.M., J.E. Keith, L. Liu and D. Blahna (2010). The Welfare Effects of Restricting Off-Highway Vehicle Access to Public Lands. *Agricultural and Resource Economics Review*, 39: 89-100.

Jones, C.A. and F. Lupi (2000). The Effect of Modeling Substitute Activities on Recreational Benefit Estimates. *Marine Resource Economics* 14: 357-274.

Kikuchi, H (1986). Segmenting Michigan's Sport Fishing Market: Evaluation of Two Approaches. PhD Dissertation, Department of Agricultural Economics, Michigan State University.

Liu, X., M.E. Lovely and J. Ondrich (2011). The Location Decisions of Foreign Investors in China: Untangling the Effect of Wages Using a Control Function Approach. *The Review of Economics and Statistics* 92:160-166.

Moeltner, K. and R. von Haefen (2011). Microeconometric Strategies for Dealing with Unobservables and Endogenous Variables in Recreation Demand Models. *American Review of Resource Economics* 3: 375-296.

Morey, E. and D.M. Waldman (1998). Measurement Error in Recreation Demand Models: The Joint Estimation of Participation, Site Choice, and Site Characteristics. *Journal of Environmental Economics and Management* 35: 262-276.

Murdock, J. (2006). Handling Unobserved Site Characteristics in Random Utility Models of Recreation Demand. *Journal of Environmental Economics and Management* 51: 1-25.

Parsons, G.R. (1991). A Note on Choice of Residential Location in Travel Cost Demand Models. *Land Economics* 67: 360-364.

Petrin, A. and K. Train (2006). Control Function Corrections for Unobserved Factors in Differentiated Product Models. Paper presented at Society for Economic Dynamics 2006 Meeting.

Petrin, A. and K. Train (2010). A Control Function Approach to Endogeneity in Consumer Choice Models. *Journal of Marketing Research* 47: 3-13.

Phaneuf, D.J., J.C. Carbone and J.A. Herriges (2009). Non-Price Equilibria for Non-Marketed Goods. *Journal of Environmental Economics and Management* 57: 45-64.

Phaneuf, D.J., V.K. Smith, R.B. Palmquist and J.C. Pope (2008). Integrating Property Value and Local Recreation Models to Value Ecosystem Services in Urban Watersheds. *Land Economics* 84: 361-381.

Randall, A. (1994). A Difficulty with the Travel Cost Method. *Land Economics* 70: 88-96.

Ricker-Gilbert, J., T.S. Jayne and E. Chirwa (2011). Subsidies and Crowding Out: A Double-Hurdle Model of Fertilizer Demand in Malawi. *American Journal of Agricultural Economics* 93: 26-42.

Rivers, D. and Q.H. Vuong (1988). Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. *Journal of Econometrics* 39: 347-366.

Simoes, J.C. (2009). Recreational Angler Surveys: Their Role and Importance Nationally and the 2008 Michigan Angler Survey. Master's Thesis, Department of Fisheries and Wildlife, Michigan State University.

Smith, R.J. and R.W. Blundell (1986). An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply. *Econometrica* 54: 679-685.

Train, K., D. McFadden and R. Johnson (2000). Discussion of Morey and Waldman's "Measurement Error in Recreation Demand Models". *Journal of Environmental Economics and Management* 40: 76-81.

Timar, L. and D.J. Phaneuf (2009). Modeling the Human-Induced Spread of an Aquatic Invader: The Case of the Zebra Mussel. *Ecological Economics* 68: 3060-3071.

Von Haefen, R.H. (2003). Incorporating Observed Choice into the Construction of Welfare Measures from Random Utility Models. *Journal of Environmental and Economic Management* 45: 145-165.

Von Haefen, R.H. and D.J. Phaneuf (2008). Identifying Demand Parameters in the Presence of Unobservables: A Combined Revealed and Stated Preference Approach. *Journal of Environmental Economics and Management* 56: 19-32.

Weber, M.A., P. Mozumder and R.P. Berrens (2012). Accounting for Unobserved Time-Varying Quality in Recreation Demand: An Application to a Sonoran Desert Wilderness. *Water Resources Research* 48.

Weithman, A.S. and M.A. Haas (1982). Socioeconomic Value of the Trout Fishers in Lake Taneycomo, Missouri. *Transactions of the American Fisheries Society* 111: 223-230.

8 Tables

Table 1. Variable descriptions

Name	Description	Mean	SD
Site Characteristics - Catch Rates			
CR chinook	Chinook (king) salmon per hour catch rate	0.028	0.055
CR coho	Coho salmon per hour catch rate	0.007	0.018
CR lake trout	Lake trout per hour catch rate	0.028	0.108
CR steelhead	Steelhead (rainbow trout) per hour catch rate	0.008	0.018
CR walleye	Walleye per hour catch rate	0.042	0.093
CR yellow perch	Yellow perch per hour catch rate	0.383	0.702
Site Characteristics - Other			
Urban	Site partially urban as defined by U.S. Census 2000	0.600	0.490
Highway	Adjacent to a state highway	0.945	0.229
Bayorseaway	Adjacent to a major Great Lakes bay or seaway	0.434	0.496
Warmwater	Warm water alternative of the site	0.500	0.500
Angler Characteristics			
Distance	Distance, in miles, to fishing site	125.113	55.112
Crude price	Crude price, in dollars/barrel, in month of trip	91.095	29.304
Travel cost	Mileage and opportunity cost of trip, in dollars	171.193	87.879

Table 2. Site choice model parameter estimates

Variable	Traditional CL Estimate	Traditional CL t-stat	CL – Only visited alternatives Estimate	CL – Only visited alternatives t-stat	CL with ASCs[a] Estimate	CL with ASCs[a] t-stat	CL with CF Estimate	CL with CF t-stat[b]
Travel cost	*-0.0348	-46.10	*-0.0345	-45.70	*-0.0359	-40.97	*-0.0357	-28.92
CR chinook	*11.759	14.10	*10.747	12.70	*7.996	4.09	*12.018	13.43
CR coho	*6.546	3.34	*6.236	3.19	6.802	2.05	*6.289	3.61
CR lake trout	0.327	0.62	-0.520	-0.88	-0.538	-0.45	0.338	0.80
CR steelhead	6.657	2.36	6.610	2.30	-4.835	-1.00	6.229	2.31
CR walleye	*3.927	12.62	*3.533	11.31	*1.915	4.26	*3.954	13.07
CR yellow perch	*0.318	6.79	*0.289	6.21	0.167	1.88	*0.303	7.15
Urban	*0.716	9.36	*0.642	8.47			*0.727	10.10
Highway	*1.801	6.88	*1.621	6.06			*1.799	6.60
Bayorseaway	*-1.091	-9.31	*-1.155	-9.92			*-1.134	-9.08
Bayorseaway Warmwater*	*1.748	15.24	*1.870	16.19			*1.753	14.71
Control Function							*0.011	3.26
Trips		2233		2233		2233		2233
Log likelihood		-4184.530		-4091.505		-3645.443		-4174.000

[a] Alternative specific constant estimates are withheld for brevity.

[b] Derived from standard errors calculated from bootstrapping with 250 replications.

*Significant at the 1 percent level.

Table 3. ASCs second-step regression

Variables	Estimate	t-stat
Urban	-0.469	-1.69
Highway	-0.298	-0.67
Bayorseaway	-0.157	-0.47
*Bayorseaway*Warmwater*	*1.260	2.94
Constant	-0.917	-2.11
R^2		0.135

*Significant at the 1 percent level

Table 4. Travel cost equation parameter estimates

Variables[a]	Estimate	t-stat
CR chinook	-3.354	-0.97
CR coho	-17.262	-1.73
CR lake trout	-1.542	-1.08
CR steelhead	20.799	2.02
CR walleye	2.573	1.28
CR yellow perch	*-1.482	-5.94
Urban	0.272	0.84
Highway	0.091	0.13
Bayorseaway	*-4.617	-11.19
*Bayorseaway*Warmwater*	0.623	1.13
January08	*-24.801	-9.42
February08	*-20.931	-7.72
March08	*-16.965	-6.10
April08	*-25.238	-9.77
May08	*-50.019	-26.93
June08	*-62.161	-36.43
July08	*-60.685	-36.55
August08	*-40.732	-25.08
September08	*-15.863	-9.82
October08	*13.310	8.05
November08	*37.803	19.87
December08	*58.344	28.41
January09	*62.999	34.75
February09	*65.689	37.47
March09	*40.851	21.25
April09	*41.156	23.01
May09	*39.095	22.53
June09	*20.687	12.13
July09	*22.582	11.75
August09	*14.431	4.57
September09	*35.483	13.78
October09	*67.224	11.12
*Crudeprice*Distance*	*0.014	476.88
Constant	*25.978	14.93
R^2		0.709

[a] The omitted category in the set of month-year dummy variables is all trips taken prior to January, 2008.
*Significant at the 1 percent level

Table 5. Average per trip WTP ($) among all Michigan Great Lakes trips for catch rate change

Scenario	Traditional CL	CL – Only visited alternatives	CL with ASCs	CL with CF
CR chinook 50% decline	3.34	3.13	2.14	3.32
CR coho 50% decline	0.50	0.48	0.44	0.47
CR walleye 50% decline	7.08	6.50	4.50	6.95
CR yellow perch 50% decline	3.25	3.00	1.62	3.03

Table 6. Average per trip WTP ($) among trips taken to a Great Lake to forgo loss of access to that lake

Scenario	Traditional CL	CL – Only visited alternatives	CL with ASCs	CL with CF
Loss of Lake Erie	27.51	27.08	39.57	26.58
Loss of Lake Huron	76.86	80.02	90.54	74.97
Loss of Lake Michigan	165.56	165.07	151.38	163.31
Loss of St. Clair and Detroit Rivers and Lake St. Clair	67.37	68.11	56.53	65.68