Calibrating Online Survey Sample for Economic Impact Analysis

Nirmala Devkota, Graduate Student
Department of Agricultural Economics and Agribusiness
Louisiana State University
Baton Rouge, LA 70803
ndevko1@lsu.edu

Krishna P. Paudel, Assistant Professor
Department of Agricultural Economics and Agribusiness
Louisiana State University
Baton Rouge, LA 70803
kpaudel@agcenter.lsu.edu

J. Matthew. Fannin, Assistant Professor
Department of Agricultural Economics and Agribusiness
Louisiana State University
Baton Rouge, LA 70803
mfannin@agcenter.lsu.edu

Larry M. Hall, Graduate Student
Department of Agricultural Economics and Agribusiness
Louisiana State University
Baton Rouge, LA 70803
lhall2@lsu.edu

Rex H. Caffey, Associate Professor
Department of Agricultural Economics and Agribusiness
Louisiana State University
Baton Rouge, LA 70803
rcaffey@agctr.lsu.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meetings Mobile, Alabama, February 4-7, 2007

Abstract

Online surveys have emerged as low-cost data collection approach in empirical researches; however, the validity of data remains questionable. Therefore, we employed a stochastic frontier estimation method to calibrate online recreational expenditure. Study results suggested the presence of inefficiency on online surveys. Analysis was extended to estimate economic impact of nature based recreation on a local economy.

Keywords: calibration, online survey, onsite survey, stochastic frontier approach

Introduction

In non-market environmental valuation researches, questionnaire survey remains the dominant approach to translate visitor's information into monetary value of natural resources. Face to face interviews and mail surveys have been the dominant methods to gather information for environmental valuation studies. Face to face data collection method incurs an extremely high cost while the mail survey is connected with a very low response rate. Further, a well designed mail survey with telephone follow up some times costs as much as conducting a face to face interview. With the advancement of the technology and internet use, online survey has emerged as potential low cost alternative data collection approach in empirical researches. However, the reliability of the estimates obtained from the online survey information still remains questionable.

A fairly recent study conducted to evaluate the stated willingness to pay (WTP) for organic fruit and pesticide ban used online and face to face interview data for the analysis (Canavari et al.). The responses obtained from conventional face to face interview were significantly smaller than those obtained from online survey. So, the authors cautioned a need to calibrate values obtained from a web-based survey before using those values for further empirical studies.

One of the previous studies by Devkota et al. estimated economic impact of recreational visitation using both the online and onsite survey data. Their study also found the estimates obtained from online survey deviated significantly from those obtained from onsite survey sample. Similar to the study conducted by Canavari et al.

their study also suggested calibrating the online survey sample before extrapolating the estimation result from online survey to the large population.

This motivation of this study is to overcome the above mentioned concerns associated with online survey by using a calibration method. The study employs the data obtained from internet survey to understand the direct and indirect effects of nature based recreation on local economic sectors. The main goal of this study is to evaluate whether online surveys can replace high cost face to face survey and low response mail survey. We employ stochastic frontier model to estimate and calibrate the true values of recreational expenditures. We then use the calibrated data to estimate the economic impact of recreational visitation on local economy in Louisiana. Thus, the objective of our study is of two fold; at first we adjust the survey information using stochastic frontier approach in order to addresses the issue of reliability of the estimates. The calibrated information is then used to estimate the economic impact of recreational visitation on a local economy.

Calibration of hypothetical values has been popular approach on contingent valuation studies to adjust the hypothetical bias associated with mail survey. Examples of such uses can be seen on List and Shogren; Hofler and List; and Fox et al. Using stochastic frontier approach Hofler and List calibrate the hypothetical willingness to pay value in order to obtain the true value of a natural resource. In this study, we employ their idea of calibration to adjust the online values obtained for recreational expenditures.

The expectation from this study is that that the online survey can provide reliable estimates if the values are calibrated using a proper method. Thus, the online survey can

replace the high cost face to face interview method employed in environmental valuation researches. In addition, our study results suggested environmental valuation studies to consider direct and indirect effects of visitor's spending on local economic sectors while providing numerical value to the natural resource.

The remainder of this paper is structured as follows. We begin our study with the description of the data collection approaches employed in this study. Then we discuss about stochastic frontier estimation techniques used to obtain the actual value of recreational cost. We further described the impact estimation procedure in the same section. Finally, we present the results, discuss the major finding, and provide conclusion and implication of the study.

Data

The stochastic frontier estimation requires the recreation related expenditure data and individual's demographic information. Similarly, the estimation of direct and indirect economic effects of recreational visitation in the regional economy requires detailed information pertaining to the out of pocket expenditure for each individual in the sample. The expenditure data for individuals traveling to the coastal Louisiana was collected using intercept and internet survey with a preset questionnaire.

Web based survey was conducted by posting the questionnaire on university's (Louisiana State University) website. The survey questionnaire remained on the web for seventy seven days starting form May 15th to July 31st, 2003. Most of the observations (approximately 92%) are obtained from online survey. Online survey was formatted in such a way that the responses were recorded in a Microsoft excel spreadsheet

automatically, once submitted by a respondent. Duplicate responses were identified and deleted for any submissions with same internet protocol address. Solicitation for the responses and announcement were made through different media including mails, radio advertisements, newspapers, magazines, websites and newsletters.

On the other hand, the intercept survey was conducted in the proxy sites the Grand Isle, LA and the Holley Beach, LA. The randomly selected individuals exiting from the recreational site were requested to fill out the questionnaire. Multiple visits to the sites were made to gather recreation related individual information. Slightly over 8% observations were gathered using onsite survey method. The total of 2691 responses was obtained using both survey methods. Several observations with no information on all of the variables used here are dropped from the data set.

The set of questionnaire contained a large number of information on individuals' and recreational site's characteristics, including average cost incurred in recreation and related activities in Coastal Louisiana. Individual expenditure section of the questionnaire is used to evaluate the estimates of economic impact calculated using web based survey information. The expenditure section provided information related to expenditures on 1) trip related supply and equipment, 2) cost incurred on the during the recreational activities in the site such as food, supplies and hotel etc and, 3) dollars spent on two way trip such as fuel.

The expenditure data included the price paid by individual for recreational and non recreational activities during a particular recreational trip. The variables include cost associated with lodging, food, fuel, parking and launching, groceries, supplies, and

equipments. Other explanatory variables include age, income, gender etc. The demographic characteristics of individuals were used as explanatory variables to predict the actual value of recreational expenditure using the stochastic frontier estimation process. whereas, the estimation of economic impact of having a recreational site in the locality uses the predicted value of spending along with input output model to obtain the impact estimates.

Methods

The expenditure associated with recreational visit is employed to estimate economic impact of recreational visitation in a local economy. Using the information obtained from the surveys and the stochastic frontier model, we predict the frontier expenditure. The predicted cost is then combined with observed cost in order to develop the calibration ratio. The calibration ratio is then employed to estimate the actual recreational expenditure. The estimated actual values ate then combined with input output model to estimate the economic impact of recreational expenditure on the local economy.

Estimation of recreational cost

We estimate the actual value of recreational expenditure using a stochastic frontier estimation technique on the web-based survey data. Stochastic frontier approach has recently gained popularity among the researches other than production and allocation efficiency. An example of such deviation includes the article by Hofler and List which estimated real value of willingness to pay using hypothetical values. We follow their concept to calibrate online and onsite survey information.

Given that the survey data are cross-sectional we define stochastic frontier recreational expenditure as;

$$Y_i^l = X_i\beta + \varepsilon_i \qquad (1)$$

Where, Y_i^l represents the recreational expenditure for an individual i captured on online survey. X_i is the column vectors of the independent variables that determine the amount of recreational expenditure for ith individual. β is the vector of regression parameters to be estimated by the model. The composite error term ε_i is made up of two independent components $\varepsilon_i = v_i - u_i$. $v_i \sim N(0, \sigma_v^2)$ is a normally distributed random error representing the usual statistical noise.

The u_i captures the distance of expenditure value from its frontier estimated using the stochastic frontier model $X_i\beta + v_i$. The u_i is assumed to have expected value $E(u_i) = \mu$ and variance $Var(u_i) = \sigma_u^2$. Thus, the error u represents a distance of online values from its actual values. This implies that the u_i represents the gap between the frontier and online survey information for an individual i. Therefore, as the value of u_i increases the gap between the online survey information and actual value of expenditure increases. And there will be no differences between the online and frontier expenditure as u_i approaches to zero. Zero value of u_i indicates there is no inefficiency term associated with online survey information. Then, the model to be estimated is expressed as;

$$Y_i^l = X_i\beta + v_i + u_i \qquad (2)$$

Where, the term $X_i\beta + v_i = Y_i^F$ represents the frontier unobserved recreational expenditure which can be estimated using individual specific characteristics.

Estimation of economic impact

At the second stage, we use the calibrated online survey data to estimate economic impacts associated with the recreational use of the site. Input Output (I-O) model has been widely used tool to evaluate the regional economic impact of developmental project, tourism industries and policy changes. Recently, input-output model has also been used to estimate economic impacts of recreational visits in regional, state and national level economy (Cordell *et al.,* Bergstrom *et al.*, English & Bergstrom, English, Lee and Choi). The model explains the estimated monetary transactions of an economy within a given period of time. It provides policy makers a view of economic interdependency existing in the economy (Henry and Deane).The main goal of the I-O model is to evaluate economic impacts of new final demand change on producing sectors in a local economy (Weiler). Impact analysis provides information on economic interdependencies of diverse industries in region's economic sectors. More specifically, it shows how changes in one sector of an economy affect all the economic activities in the region, state or in a nation.

A standard input output model which is used to obtain the output multipliers is expressed as;

$$\mathbf{q = [I - A]^{-1}f} \qquad (3)$$

Where **[I - A]** is a representation of Leontief inverse matrix that translates a particular level of final demand into direct and indirect outputs from each sector of region's economy required to meet the final demand (**f**).

In general, input-output model holds a number of assumptions. First, it assumes an economy consists of N number of sectors each producing one commodity, and a final demand sector. Second, the firms show a constant return to scale such that there are no external economies or diseconomies. Third, firms have no supply constraints to meet the increased demand. Fourth, there is a linear dependence between inputs and level of outputs in an economy. And finally, there is no substitution of intermediate inputs used. Despite of these binding assumptions, I-O model has been widely used in tourism literatures because of difficulty in finding any other more reliable and appropriate tools.

Results and Discussions

The economic impact of recreation on local economy is estimated at two steps. First, we estimated stochastic frontier model on online survey data to estimate the actual value of recreational expenditure. For the estimation, we used the recreational expenditure obtained form online survey as dependent variables and income, age, gender and whether an individual has a full time job as explanatory variables.

The average recreational spending per individual using internet and onsite surveys are presented in Table 1. The recreational expenditure pre individual varied from $11 for parking to $101 for lodging in case of onsite data. While, the expenditure varied from about $10 for parking to $56 for lodging in case of online data. Online data showed

significantly smaller expenditure values than those came from onsite survey indicating some concern over the data. Therefore, using the values directly from the online survey would underestimate the true value of a natural resource.

We therefore, tested the online data for the presence of inefficiency associated with it. The result showed that the inefficiency term u_i is significantly different from zero causing some level of variation on the online data due to presence of one sided error. Onsite survey data was also tested to check whether there is inefficiency in onsite data. The result showed that the data obtained form onsite face to face interview showed the absence of inefficiency in the data set. The one sided error term u_i was not significantly different from zero. So, stochastic frontier model suggested the onsite survey data did not contain the variation due to one sided error term.

The stochastic frontier model was estimated using online expenditures as dependent variable and all the explanatory variables. We estimated the log linear form of the stochastic frontier recreational expenditure function. Most of the explanatory variables are contributing significantly to the predicted value of recreational expenditure for all categories of the cost.

Once the frontier expenditure was predicted the calibration ratio was calculated using the predicted and observed values of the expenditure. The calibration factor is defined as the ratio of mean frontier expenditure to the mean online expenditure data. The online data is then adjusted using the gap ratio to obtain calibrated value of online survey data.

Each individual's recreational expenditure was adjusted by using the gap ratio to get the actual value of expenditure. For example if an individual has spent 50 dollars on food and the calibration ratio is 1.6 then the person's true expenditure is 80 dollars. This method is equivalent to one of the two approaches taken by Hofler and List to calibrate hypothetical value of willingness to pay for sport card to calculate the actual value. Their study suggested that the approach we have applied here provided more conservative and accurate calibrated values than the other method in their study.

The calibrated values of the expenditure were then used in the Impact Analysis tool (IMPLAN). The IMPLAN model now derives a regional version of the Input Output model by using county level data. IMPLAN doesn't define a specific sector as "tourism" within its default set of 509 economic sectors. To address this issue, we used MI-REC spreadsheet which consists of a set of utilities and customized procedures for estimating the economic impact of recreational and tourism spending. Mean recreational expenditure is combined with the IMPLAN MI-REC bridge table to calculate estimated impacts on the local economy. MI-REC contains eleven sectors of recreation-related expenditures which are bridged to 509 IMPLAN sectors.

The table 2 shows the effect of dollars spent recreational purposes on the coastal Louisiana. The output effects are categorized according to two digit NAIC codes. The result showed that the most benefiting sector of local economy was art and entertainment sectors which are directly related to the recreation. Retail sectors and the manufacturing industries are also benefiting from the recreational visitors more than other sectors of the economy. On an average, the total economic effect of an individual's recreational

expenditure on the local economy varied from $287 to $643 based on the estimation procedure and survey mode. The result showed that there are other sectors in the economy which are also directly and indirectly benefited by the use of natural resource for recreation. Our result implied that the value of a natural resource is more than just an individual's wiliness to pay for the pleasure he gets by using the resource for recreation or just the cost incurred on a particular recreational trip. The direct and indirect effect of visitor's spending on local economy is also an important factor to be considered while estimating the economic value of a natural resource.

Implications

This study has an implication for the researchers intending to use online survey data. Our study result indicated the presence of inefficiency in the online data causing the variation in the data other than those contributed by the normally distributed random errors. Self selection in the online sample is the obvious factor and probably the potential reason behind the fact that the distribution of online survey data significantly differ from that of onsite survey data. It suggested searching for some *ex post* correction procedure to obtain reliable estimates from the online survey data.

In addition, the economic impact estimation results would be helpful for the planners and policy makers performing economic impact assessments of some project that may affect the nature based recreation adversely. The quantitative information obtained from this research will help policy makers to understand the numeric value of those particular natural resources.

It would also be helpful to understand the importance of a natural recreational site to the local economy. The result will add to make more educated decision on how to manage and preserve such a valuable resource for recreation in more economically and environmentally efficient manner. However, the result implies that the researchers should be careful in using web-based data.

Conclusion

We calibrated the online data to estimate economic impact of recreational site on a local economy and found that the estimation using online survey under predicted the true value of a natural resource. Using stochastic frontier approach in combination with the IMPLAN our study estimated the effect of recreational spending in Louisiana's different economic sectors.

Using online survey sample we estimated the frontier value of recreational expenditure. The calibration ratio using online and predicted values is then estimated and used to calibrate the online survey values. The calibrated sample is then employed to estimate economic impact on input output model. At this preliminary phase of our study the calibrated values are still different than that of onsite values. Model specification might be the potential reason behind this fact.

References

Bergstrom, J, C., H. K Cordell, A. E. Wateson, & G. A. Ashley. Economic impacts of state parks on state economics in the south. *Southern Journal of Agricultural Economics,* 29 (1990): 69-77.

Canavari, M., G. Nocella, R. Scarpa. Stated willingness to pay for organic fruit and pesticide ban: An evaluation using both web-based and face to face interviewing. *Journal of food Products Marketing* 11(3) (2005): 107-134

Cordell, H. K., J. C. Bergstrom, A. A. Gregory, and J. Karish. Economic effects of river recreation on local economies. *Water Resource Bulletin* 26(1) (1990) :53-60.

Devkota, N., J.M. Fannin, and K. Paudel. Economic impact estimation using bootstrap samples obtained form internet and intercept survey data. Selected paper at the American agricultural Economics Association Annual Meeting, Long Beach, California, July, 23-26, 2006.

English, D. B. K. Calculating confidence interval for regional economic impacts of recreation by bootstrapping visitor expenditure. *Journal of Regional Science,* 40(3), (2000): 523-539

Fox, J.A., J. Shogren, D., Hayes, and J. Kliebenstein. CVM-X: Caligrating continsgent values with experimental auction markets. *American Journal of Agricultural Economics.* 80(August 1980):455-465.

Henry, E.W. and B. Deane. The Contribution of Tourism to the Economy of Ireland in 1990 and 1995. *Tourism Management* 18 (1997): 535-553.

Hofler, R.A., J. List. Valuation on the frontier: calibrating actual and hypothetical statements of value. *American Journal of Agricultural Economics.* 86(February 2004): 213-221.

Lee, K. and S. Choi. An Economic Impact Analysis of Regional Tourism Industry the Case Stury of Gangwon and Jeju regions *Proceedings of the New Zeland Tourism and Hospitality Research Conference* (2004): 199-210.

List, J. and J. Shogren. Calibration of difference between actual and hypothetical valuations in a field experiment. *Journal of Economic Behavior an Organization.* 37(1998): 193-205.

Weiler, S. and A. Seidl. What's in a Name? Extracting econometric drivers to assess the impact of national park designation. *Journal of Regional Science,* 44 (2) (2004): 245-262, 2004

Appendix:

Table 1: Average recreational expenditure based from different source

Variables	Survey	Predicted	Online
Lodging	101.17	212.04	56.10
Fuel	38.97	65.58	44.71
Food and beverages	70.43	49.61	45.16
Equipments	89.81	61.59	28.37
Supplies	37.95	49.53	26.47
Parking and Launching	11.37	19.23	9.95
Miscellaneous	62.73	81.52	25.44

Table 2: Estimated economic impact on local economy

Variable name (IMPLAN sector category)	Output effect for calibrated online	Output effect for original onsite	Output effect for un-calibrated online
Agriculture and forestry	0.777	0.687	0.444
Mining	24.225	14.615	16.019
Utilities	4.202	2.997	1.906
Construction	3.884	2.613	1.495
Manufacturing	102.86	71.902	55.718
Wholesale	40.092	34.615	20.319
Transportation	13.457	10.02	6.573
Retail industry	90.563	85.613	40.639
Information	19.089	14.589	6.659
Finance	6.35	4.833	2.834
Real estate	22.961	16.434	10.218
Technical	18.266	13.346	7.731
Management	0.593	0.446	0.252
Administration	8.721	6.099	3.437
Education	0.568	0.44	0.247
Healthcare	0.168	0.221	0.142
Art and entertainment	207.183	99.377	55.139
Hotel business	57.751	74.429	47.38
Restaurant	0.079	0.069	0.037
Others	21.841	14.045	10.749
Total	643.628	467.389	287.938