Þjóðarspegillinn 2010

Valuing recreational demand

The case of Heiðmörk

Daði Már Kristófersson
Kristín Eiríksdóttir

Hagfræðideild
Ritstjóri: Daði Már Kristófersson

Rannsóknir í félagsvísindum XI. Erindi flutt á ráðstefnu í október 2010
Ritstýrð grein
Reykjavík: Félagsvísindastofnun Háskóla Íslands
ISBN 978-9935-424-04-4

HÁSKÓLI ÍSLANDS

Valuing recreational demand

The case of Heiðmörk

Daði Már Kristófersson
Kristín Eiríksdóttir

To this day the full social cost of proposed projects in Iceland has not been taken into account in the decision making process since environmental goods being sacrificed have not been valued in monetary terms. Therefore, the monetary impact of socially desirable goods, such as unspoiled nature and access to outdoor recreational areas, is unknown. This paper is a part of a comprehensive environmental valuation study of the largest recreational area in the greater Reykjavik area, Heidmork. The project is funded by Rannis, the city of Reykjavik, the municipality of Gardabaer and Reykjavik Energy and its primary objective is to improve the current situation in Iceland and serve as a point of reference for future reaserch of valuable Icelandic environmental goods and services.

This paper focuses on the attributes of data gathered on recreational behavior in Heidmork in the years of 2008 to 2010 and the apparent problems of immensly heterogenous user groups, the existence of heavy users and to a large extent a lack of variance in the indiviual travel costs. When left unattended, any one of those problems could cause a conventional recreational demand estimation with a count data model specification to break down or to at least give biased welfare estimates. This paper proposes a comprehensive plan that combines revealed and stated preference data, calibrations of recreational demand with real time traffic counts and finally by estimating a recreational demand model with a latent-class negative binomial model to address the above mentioned problems. Such an approch can potentially shed light on and handle both the observed and unobserved factors of user group heterogeneity, the thick upper tail problem of heavy users and the lack of distribution in individual travel costs.

Advances in the Travel Cost Method

Intuitively, the use of a recreational site and the trip to the site are complimentary goods and therefore travel cost can serve as an implicit price for recreation. The travel cost method (TCM), simplicitically described above, dates back to Hotelling who laid down the foundation of the method in his 1947 letter to the National Park Service (Phaneuf and Smith, 2005). Ever since, the TCM has proven to be a powerful tool to estimate recreational demand and its welfare implications. Early applications focused on single site count data travel cost models to estimate recreational demand but as advances were made in discrete choice theory with McFadden's random utility model (RUM) (see McFadden, 1974) the TCM literature gradually shifted towards site choice models with Hanemann's disseration (see Hanemann, 1978) setting the stage.

Applications and the theoretical foundation of the single site count data travel cost model are well documented in Shaw (1988), Creel and Loomis (1990), Haab and McConnell (2002), Parsons (2003) and Herriges and Kling (2003). The endogenously stratified, zero-truncated, negative binomial model is the predominant estimation

model for travel cost count data gathered on-site due to the tendancy of data to be overdisperse (see Martinez-Cruz, 2010). For a theoretical treatment the reader is referred to Cameron and Trivedi (1998), Haab and McConnell (2002) and Hilbe (2007).

Advances in the site choice model have on the one hand focused on fitting models with ever better prediction power and on the other hand on ways to improve the quality of datasets by combining revealed and stated preference data. Site choice models have evolved from the multinomial logit RUM model, first applied by McFadden et al. (1977) on transport choice data, to more complicated mixed logit models that can better model unobserved heterogeneity and substitution patterns with different mixing distributions for the underlying parameters (see Hensher, Rose, &, Greene, 2005; Train, 1999, 2009). Recent empirical work in the nonmarket valuation literature focuses on the use end evidence in the support of latent-class models (see Scarpa & Thiene, 2005; Train, 2008), also called finite mixture models. Latent-class models take advantage of the EM algorithm, first proposed by Dempster, Laird and Rubin (1977) for handling missing data, to estimate flexible distributions of preferences which is undeniably convenient in the presence of immense heterogeneity. Applying latent-class models to recreational demand data has tremendous policy implications since the EM algorithm is capable of mapping the prefernces of different user groups by dealing with inter- and intra-class heterogeneity simultaniously. The reader is reffered to McLachlan and Krishnan (1997) and Train (2009) for theoretical treatment and applications of the EM algorithm.

Combining revealed and stated preference datasets is a way of data enrichment (see Louviere, Hensher, & Swait, 2000) and gaining more information on relative weights of different user groups and observed heterogeneity accross users than would be possible with an on-site sample alone. To combine datasets the researcher imposes the restriction of equlity of parameters of common attributes across datasets given the difference in scaling. For a site choice model the problem of different scales can be sovled with a nested logit model. For readings on combining stated and revealed preference data see Ben-Akvia and Morikawa (1990), Adamowitcz, Louviere and Williams (1994), Louviere et al. (2000) and Haab and McConnell (2002).

Recently, the focus of the travel cost literature has been turned back on single site estimation with latent-class count data estimation with applications by Scarpa, Thiene and Tempesta (2007) and Martinez-Cruz (2010). These applications suggest evidence in support of estimating single site recreational demand with finite mixture models.

Data

Heidmork is an open access urban park owned by the city of Reykjavik, the municipality of Gardabaer and Reykjavik Energy. It is by far the largest recreational area in the vicinity of the greater Reykjavík area and is used by diverse types of user groups year round. The park encompasses around 3200 hectares of forests, lava fields, lakes, a water basin and open areas as well as over 40 kilometers of walk paths. Some of the different receational activities the park has to offer include walking/hiking, running, biking, horseback riding, cross-country skiing, fishing, picnicing, picking berries and mushrooms and nature watching as well as being popular amongst residents of the capital area for going for a drive. Two datasets were gathered along with hourly traffic counts from the entry and exit points of the park. The user groups are observably heterogenous with respect to different recreational activities, different seasons of the year, residency and socio-econmic standing such as age, family size, job participation and income.

On-site survey

Several pilots were carried out on-site in Heidmork in June of 2008. Their objective was twofold: Firstly, the questionnair in its entirety was tested as well as phrasing and ordering effects of individual questions. Secondly, the pilots were used to determine the optimal locations for data gathering in terms of cost effectiveness and to maximize the participation rate in whole and across different user groups. Both the entry and exit points to Heidmork and the vast number of parking areas posed a problem for the sampling methodology. Heidmork can be reached by foot, bycycle or horse in numerous locations but there are only two points of entry and exit available for vehicles to the main recreational area. There are dozens of parking areas in Heidmork and two of those are located on the outskirts of the park.

With the seasonal aspects of recreation in Heidmork and the heterogenous distribution of users within the park in mind data from the users was gathered through a self administrated survey from the beginning of July of 2008 until the end of September 2009. Survey days were chosen at random with 3-4 survey days a month during the summer months, May through September, and 1-2 survey days a month during rest of the year. Within each survey day a 4 hour continuous survey period was chosen at random given the constraint of daylight hours. Users were surveyed at both of the entry and exit points for traffic and at the two parking areas at the outskirts of Heidmork when they were leaving the park and they were asked to report, among other things, the number of trips taken to the park in the last calander month and their street adress and zip code. The process resulted in approximately 2500 observations and an overall participation rate of 67%. The roundtrip distance in kilometers and roundtrip travel time in minutes was measured with a traffic GPS system provided by www.ja.is. By analysing some of the descriptive statistics from the dataset a lack of variation in the travel cost and the problem of heavy users becomes apparant, e.g.:

- 99% of participants reside in the capital area (including Reykjanes),
- the average roundtrip distance for participants reciding in the capital area is 18.8 kilometers and the standard deviation is 10.52 kilometers,
- the standard deviation of roundtrip distance within each zip code is in most cases around 1 kilometer, which is well within the range that could usually be attributed to mere measurement error,
- 73% of participants take 4 trips, the average, or fewer,
- 90% of participants take 10 trips or fewer,
- 2.2% of participants take 25 trips or more per month,
- 44% of participants report other recreational activities besides the core activities of walking/hiking, running, biking, going for a drive and horseback riding.

Online survey

The sampling methodology of the on-site survey failed to capture certain user groups such as horseback riders, runners and bikers. Horseback riders in Heidmork tend to ride in groups which makes it dangerous to try and stop them. Both runners and bikers were in general reluctant to participate in the survey since often times they were still in the midst of their timed excercise routine. To gain more information on those user groups as well as to reveal the relative weights of different user groups in the underlying population, an online survey was designed.

In June of 2010 four thousand individuals, randomly sampled from the population of Iceland, were asked to participate in an online survey about Heidmork carried out by Capacent, an Icelandic market research company. The participants were unaware of

the survey topic until they were well on their way with answering the questions. Of the approximately three thousand individuals that participated in the survey 59% had visited the park in the year 2009. Even though the sampling proceedure excludes the self-selection of heavy users into the sample the descriptive statistics from the dataset reveal the same pattern of heavy users as the on-site sample did. The following are highlights from the above mentioned descriptive statistics of sampled users:

- The average number of trips taken by participants lies in the range of 6-11 times in the year 2009,
- 91% of participants visit the park on average less than once a month in the year 2009,
- 41% of participants take two or fewer trips in the year 2009,
- 2% of participants visit the park on average at least once a week the entire year of 2009,
- 29% of sampled visitors report other recreational activities besides the core activities of walking/hiking, running, biking, going for a drive and horseback riding.

Traffic Counts

To gain reliable information on total use of Heidmork, The Icelandic Road Administration implanted traffic censors into to the Asphalt of both of the entry and exit points to the park. The censors provided data that can be broken down into hourly intervals for the entire year of 2009 and can easily be cross-referenced with the travel cost data gathered on-site further aiding the mapping of heterogenous user groups with respect to travel.

Discussions

After consulting with experts in matters concerning Heidmork it was evident that in order to get representative and unbiased welfare estimates associated with recreational use of the park a straight forward on-site sample would not suffice due to seasonal aspects of the user groups as well as the difficulty of reaching certain user groups while on-site. Therefore, a dataset including an on-site sample, online survey data and hourly traffic counts in Heidmork was collected.

Since several steps were taken in the sampling methodology of the on-site sample to ensure a representative sample of heterogenous user groups, the authors are lead to believe that the problem of heavy users cannot be solved by endogenously stratifing the model but rather is explained by the nature of the urban park in question. In other words, we believe that the overdispersion in the data is due to unobserved heterogeneity of users rather than statistical properties of data gathering process. This belive is reinforced with the online survey data where it is noteworthy to mention that 91% of sampled users report a number of trips taken in the year 2009 that lies in the same range as the average reported trips. Furthermore, to solve estimation problems likely to arise due to the lack in variation of travel costs accross individual users of Heidmork with any measures other than a model that is well suited for preference mapping would be missing the point of welfare estimation.

By combining all the best approaches the TCM literature has to offer the problems of heterogenous user groups, existence of heavy users and lack of variation in travel costs can be addressed simultaneously by estimating a latent-class negative binomial model of recreational demand for Heidmork. The next step in this research is to formalize a latent-class count data model that will incorperate the data collected on-site as well as the online survey data. After estimation of a recreational demand model

the traffic counts will then be used to estimate the total value of recreation in Heidmork.

References

Adamowitcz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management, 26*, 271-292.

Ben-Akiva, M., & Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transport Research, 24*, 485-495.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. New York: Cambridge University Press.

Creel, M. D., & Loomis, J. B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *American Journal of Agricultural Economics, 72*, 434-441.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B 39*, 1-38.

Haab, T. C., & McConnell, K. E. (2002). *Valuing environmental and natural resources*. Cheltenham, UK: Edward Elgar.

Hanemann, W. M. (1978). *A methodological and empirical study of the recreation benefits from water quality improvement*. Ph.d. Dissertation (Economics). Harvard University.

Hensher, D. A. Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis, a primer*. Cambridge, UK: Cambridge University Press.

Herriges, J. A., & Kling. C. L. (2003). Recreation demand models. In J. Folmer & T. Tietenberg (Eds.), *International Yearbook of Environmental and Resource Economics 2003/2004* (pp. 331-372). Cheltenham, UK: Edward Elgar.

Hilbe, J. M. (2007). *Negative binomial regression*. Cambridge, UK: Cambridge University Press.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods, analysis and application*. New York: Cambridge University Press.

Phaneuf, D. J., & Smith, V. K. (2005). Recreation demand models. In K. G. Mäler & J. R. Vincent (Eds.), *Handbook of Environmental Economics* (pp. 671-761). Amdsterdam: Elsevier B.V.

Martinez-Cruz. A. L. (2010). *Handling overdispersion provoked by heavy users of natural parks*. Unpublished. Retrieved July 10, 2010, from http://www.webmeets.com/files-/papers/WCERE/2010/1530/handling_overdispersion.pdf

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

McFadden, D., Talvitie, A., Cosslett, S., Hasan, I., Johnson, M., Reid, F. et al. (1977). *Demand model estimation and validation, final report, volume V, urban travel demand forcasting project*. University of California Berkeley: Institute of Transportation Studies, Univeristy of California Berkeley.

McLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensinons*. New York: Wiley.

Parsons, G. (2003). The travel cost model. In P. A. Champ, K. J. Boyle & T. C. Brown (Eds.), *A Primer on Nonmarket Valuation* (pp. 269-329). Netherlands: Kluwer Academic Publishers.

Scarpa, R., & Thiene, M. (2005). Destination choice models for rock climbing in northeastern alps: A latent-class approach based on intensity of preference. *Land Economics, 81*(3), 426-444.

Scarpa, R., Thiene, M., & Tempesta, T. (2007). Latent class count models of total vistitation demand: Days out hiking in the eastern alps. *Environmental and Resource Economics, 38*(4), 447-460.

Shaw, D. (1988). On site samples' regression problems of non-negative integers, truncation, and edogenous stratification. *Journal of Econometrics, 37*, 211-223.

Train, K. (1999). Mixed logit for recreation demand. In J. Herriges & C. Kling, (Eds.), *Valuing Recreation and the Environment*. Northhampton, MA: Edward Elgar.

Train, K. (2008). EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling, 1*, 40-69.

Train, K. (2009). *Discrete choice methods with simulation* (2nd ed.). New York: Cambridge University Press.