Applied Economics Journal Vol. 23 No. 2 (December 2016): 48-74

ISSN 0858-9291

Received: 17 October 2016
Received in revised form: 11 January 2017
Accepted: 26 January 2017

Validity of Internet-based Stated Preference Data in Modeling Waterfall Recreation Site Choice*

Sarun Kamolthip
School of Development Economics, National Institute of Development Administration
Seri Thai Road, Klong Chan, Bangkapi, Bangkok, 10240 Thailand, E-mail: s.gun.kamolthip@gmail.com
Udomsak Seenprachawong
School of Development Economics, National Institute of Development Administration
Seri Thai Road, Klong Chan, Bangkapi, Bangkok, 10240 Thailand, E-mail: udomsak.s@nida.ac.th

Abstract

This study aims to validate the use of stated preference (SP) data collected from self-administered internet survey in modeling recreational demand. Variety of tests and measures are conducted to test whether the internet SP data yields consistent information with in-person interview SP data. The probabilistic conditional logit model is used to analyze waterfall site choices of day-trip recreationists. For preference homogeneity test, the underlying preference structure of the internet SP data is not statistically different from that observed from revealed preference (RP) data, whereas the underlying preference structure observed from the SP data – which was a part of the RP survey – is not always the same as that observed from the RP data. For predictive ability test, variety of tests and measures indicate that the in-person interview SP data is not always superior to the internet SP data. With the caveat of confounding sample frames, the findings of both tests consistently suggest that the recreational site choice models that use in-person interview SP data are not always superior to the models that use internet SP data. Our findings do not support what is often assumed that the SP survey which administers in-person would provide superior data quality. The study indicates a great potential of internet survey as an alternative survey mode for the hypothetical study of recreational demand.

Keywords: Internet survey, validity test, stated preference, recreation site choice
JEL Codes: C83, Q26, Q51

*This article is part of Kamolthip's Ph.D. dissertation for the School of Development Economics at the National Institute of Development Administration.

Background and Motivation

The US National Oceanic and Atmospheric Administration (NOAA) Panel recommends that the in-person interview is more preferred in stated preference (SP) study because it would enable the interviewer to present the visual materials such as maps and pictures in order to facilitate respondent understanding (Arrow, Solow, Portney, Leamer, Radner, & Schuman, 1993; Carson, 2000). Although the survey costs of in-person interviews in developing countries are usually lower than that in developed countries given the same sample size (Whittington, 1998), in-person interviews are relatively very expensive, and impractical to implement in many situations. Lower cost survey mode for SP study while maintain a high degree of reliability is therefore needed (Carson, 2000).

Through the last 20 years, the internet has become a global phenomenon and an important medium of communication in our society. The website of National Electronics and Computer Technology Center (NECTEC) reported that 38,015,725 of Thais had internet access in 2015, compared with 30 Thais when the internet was first introduced in 1991. The rapid growth of the internet has opened new opportunities for both academics and government officials. Its audio-visual ability to communicate information, large number of samples, and lower unit cost (compared with in-person interview, mail surveys) provides a huge potential in conducting and design of surveys. The internet survey is therefore one of the interesting alternative data collection modes for the policy makers.

The marketing and transportation researchers have long recognized advantages of the internet survey in the SP study.[1] Although the internet survey has increasingly been implemented in the area of environmental and natural resource economics, this survey mode is however not widely adopted in the SP study. Besides the complication associated with the hypothetical nature of the SP questions (Swait & Adamowicz, 2001a; 2001b), the issue of representativeness generated by the internet survey has not yet been addressed. The research on the effects of internet survey on the derived stated preference and value measures attracts great attention in the SP literature. Lindhjem & Navrud (2011b) provide a thorough review on mode comparison studies which compare the use of internet SP survey to the other competing survey modes for environmental non-market goods (see also Boyle, Morrison, MacDonald, Duncan, & Rose, 2016). With the caveat of confounding factors,[2] they conclude that most studies reviewed have fairly similar degrees of validity and general data quality between internet and the other survey modes.

---

[1] See Miller, Hofstetter, Krohmer, & Zhang (2011) and Collins, Rose, & Hess (2012) for recent works in marketing and transportation literature, respectively.

[2] Controlling confounding factors in the mode comparison studies is challenging. Except for Lindhjem & Navrud (2011a) and Boyle et al. (2016), the other studies suffer from many confounding factors (e.g. different sample frame, different timing of implementation, different in survey presentation, different sample sizes, etc.). Hence, the differences identified in most of the mode comparison studies are not clear whether they are due to the survey mode or some confounding effects.

Confounding sample composition effects seem to be one of the most difficulties in survey mode study in SP literature (Lindhjem & Navrud, 2011b). Lindhjem & Navrud (2011a) and Boyle et al. (2016) are only two studies that strictly control the sample frame when investigating differences in internet and other survey modes in the context of SP study. Lindhjem & Navrud (2011a) investigate differences in internet and face-to-face (in home) interview survey modes. They find that the use of internet in contingent valuation study do not seem to be significantly different or biased compared to face-to-face interviews. Boyle et al. (2016) investigate differences in internet and mail survey modes in the context of a choice-modeling study. They find that the mode effects of using internet survey are statistically significant and quite large. However, the effect of sample frame (internet panel vs. postal addresses) on welfare estimates for implementation of a mail survey is not significant.

Most of the mode comparison studies test the hypothesis of equivalent mean marginal willingness to pay (WTP) derived from two different SP data sets of internet and the other survey mode. Since the SP data are nevertheless not based on actual market behavior, and their validity is always an issue (Ben-Akiva, Bradley, Morikawa, Benjamin, Novak, Oppewal, & Rao, 1994), such comparisons are therefore not easily convinced. On the other hand, it is widely accepted that that the actual behavior data is superior to hypothetical behavior data when modeling demand for recreation (Ben-Akiva and Lerman, 1985, p.368). Thus, instead of directly comparing estimated preferences and the mean value estimates for the internet survey with the other survey modes (in our case, the in-person interview); this study proposes to investigate the differences in data quality generated from the internet and in-person interview surveys by comparing each single set of these SP data to the RP data. Both internal and external validity are examined so as to complete the validation. Explicitly, we assume that the actual behavior-based RP data is superior in both explaining the actual behavior and predicting the choice of recreationists to the SP data.

This study focuses on the recreational use values of *day-trips* taken to the waterfall sites around Khao Yai National Park (KYNP). Specifically, the probabilistic conditional logit model is used to analyze recreationists' site choice (Hanemann, 1984). For internal validity, with available revealed preference (RP) data for the demand on waterfall recreation (see Kamolthip, 2016), it is possible for us to combine it with the SP data for which the same choice context and attributes were customized. These combinations of the RP data and each single set of SP data then allow us to test their preference homogeneity using the procedure first proposed by Swait & Louviere (1993) (see, also, Louviere, Hensher, & Swait, 2000). Although the test of internal validity for the pooled SP-RP data is in itself the indirect test the external validity of SP data (Ben-Akiva et al., 1994), we further examines the true external validity of both data sources by testing their predictive ability (Horowitz & Louviere, 1993; Haener, Boxall, & Adamowicz, 2001). This test is deemed much stricter test in validating the SP data (Ben-Akiva et al., 1994), and, to my knowledge, none of mode comparison studies has included this validity test.

The remainder of this paper organized as follows. Next section describes stated preference choice experiment design, survey implementation, and data used in the study. The theory is briefly discussed in the third section, followed by detail of econometric specifications used in subsequent analysis. The fourth section presents preliminary results of estimations on three single data sets (RP, in-person interview, and internet data) for the purpose of coefficient interpretation. Detail of each validity test and relevant results is then discussed in the next section. Following discussion on interesting findings in the sixth section, the last section provides conclusions and an idea for future research.

Waterfall Recreation around Khao Yai National Park: the Choice Experiment and Data Collection

This study used waterfall recreation around KYNP as a case study. KYNP is one of the most popular national parks in Thailand in which a large number of recreational spots and activities are available for recreationists. One of the most popular recreational spots is waterfalls which, currently, thirty sites of reachable waterfalls have reportedly been explored. Specifically, the study focused on the recreational use values of *day-trips* taken to the assumed alternative choice set of 10 waterfalls around KYNP which were deemed suitable for day-trip recreation. These included Takro falls, Ched Sao Noi falls, Heaw Narok falls, Heaw Suwat falls, Nang Rong falls, Sa Rika falls, Pha Kluay Mai falls, Krong Kaew falls, Muak Lek falls, and E-To falls.[3]

In order to help policy makers make better informed decisions, the SP study on waterfall recreation site choice was conducted to get more understanding in recreationists' preferences for waterfalls' characteristics and park's facilities. The study was a supplement to the actual behavior RP study on the demand for waterfall recreation around KYNP. Considering the plan of the Department of National Park, Wildlife, and Plant Conservation (DNP) to develop facilities to attract more visitors (DNP, 2012), the choice experiment was highly suitable since stated preference methods could involve the new situation that was probably outside the current set of experience to model recreationists' site choice behavior (Adamowicz, Louviere, & Williams, 1994).

As part of the RP study, the stated choice data of in-person interview was from intercept survey conducted on the weekends between June and July of 2015 at six waterfalls. Detailed information about data collection and questionnaire design can be found in Kamolthip (2016). The attributes and the levels of the attributes used in the choice experiment were guided by the literature related to nature-based recreation management in Thailand. Specifically, recreation resource potential (RRP) and recreation opportunity spectrum (ROS) studies in relation to waterfall recreation (Tanakanjana, Arunpraparut, Pongpattananurak,

[3] This is due mainly to a lack of site characteristic data. Nevertheless, since the ultimate purpose is not deriving the mean value estimates, the effect of such unrealistic choice set is therefore not expected.

Nuampukdee, & Chumsangsri, 2006; Tanakanjana, Phumsathan, & Nunsong, 2012) were used as guidelines. Table 1 presents the attributes and levels used in the case study.

Table 1: List of attributes and levels used in the case study

Attributes	Description	Levels
Type of waterfall (TYPE)	Physical characteristics of the drops of the waterfall	Water descends vertically. Drop(s) is then clearly seen
		Water descends along gradually sloped surface. Drop is not clearly seen*
Water flow (FLOW)	The flow characteristic of the stream	Strong stream
		Slow stream*
Swim area [a]	Number of permissible swimmable spots	Not allowed*
		Very few (SWIM_L)
		Moderate (SWIM_M)
		Plenty (SWIM_H)
Picnic spots [a]	Number of picnic spots around the waterfall body or the connecting stream	Not allowed*
		Very few (PICN_L)
		Moderate (PICN_M)
		Plenty (PICN_H)
Entry distance (ENTRY)	Shortest walking distance from parking area to the waterfall body or its connected stream	Short (less than 500 m.)*
		Distant (more than 500 m.)
Location and the quality of surrounding nature (NATQ)	Location waterfall and the quality of surrounding nature	The waterfall is located within the national park and, thus, has a good environmental quality
		The waterfall is located within the non-national park area and, thus, has relatively lower environmental quality*
Quality of walkway (WW)	The quality of walkway inside the park	Mostly dirt, to conserve the nature of the park*
		Mostly paved, to facilitate the visitors
Natural trail (TRAIL)	Availability of natural trail inside the park	Not available*
		Available
Interpretive media (SIGN)	Numbers of interpretive media to provide knowledge to visitors	Rarely seen*
		Highly visible
Availability of Highway (HWAY)	Availability of highway in travelling to the park	Highway is available
		Only local road to this park*
Entrance fee (FEE)	Entrance fee	Free / 25 / 50 / 75 THB
Travel distance (DIST)	From visitor's residence to the park	50 / 100/ 150/ 200 kilometers

Notes: 1. * indicates a base level for effect codes.

2. Unless otherwise specified, abbreviation of each variable used in all models is shown in parentheses.

3. [a] indicates that, for RP model, this coefficient represents “availability” of associated attribute, instead of the levels.

Suppose there are ONLY two waterfalls – A falls and B falls – available for your next recreational opportunity.
Main characteristics of each waterfall are described below

	Main characteristics of each waterfall	
	A falls	**B falls**
Physical characteristics of the waterfall and the stream	• Water descends vertically. Drop(s) is then clearly seen	• Water descends along gradually sloped surface. Drop is not clearly seen
	• Strong steam	• Slow steam
	• Not allowed	• Moderate number of permissible swimmable spots
	• Plenty of picnic spots around the waterfall body or along the stream	• Not allowed
	• Distant (more than 500 m.)	• Short (less than 500 m.)
	• The waterfall is located within the non-national park area and, thus, has relatively lower environmental quality	• The waterfall is located within the national park and, thus, has a good environmental quality
Infrastructure and facilities provided in the park	• Mostly paved, to facilitate the visitors	• Mostly dirt, to conserve the nature of the park
	• Available	• Not available
	• Highly visible	• Rarely seen
	• Very convenient due to availability of the highway	• Moderately convenient. No highway available, only local road to this park
Entrance fee	• 75 THB per person	• 25 THB per person
Travel distance from your home to this waterfall	• 100 kilometers	• 150 kilometers

If you have an opportunity to take a trip to the waterfall, which waterfall will most likely be chosen?

☐ A falls ☐ B falls ☐ Stay at home

Figure 1: An example of a choice set used in the study

Shifted design approach (Bunch, Louviere, and Anderson, 1996) was used to construct an optimal generic choice experiment. The design was selected from the collective full factorial main effects design of 2^8x4^4 or 65,536 combinations. The statistical software was used to create a D-efficiency design (Kuhfeld, 2010). As a result, a set of 32 combinations which contain the highest D-efficiency (D-efficiency = 90.37) were generated. These choice sets were then blocked into eight blocks each containing four choice sets to minimize respondent's fatigue when answering the questionnaire. The respondents sampled in the study were randomly presented with one of these blocks of four choice sets. Each choice set contained a pair of alternative descriptions of waterfall sites plus a "stay at home" option. An example of a choice set is presented in Figure 1. The same set of choice sets were implemented in both in-person interview and internet surveys.

Apart from the choice experiment questions, the respondents were also asked to report the number of visits to each of pre-defined waterfall sites taken between January 2014 and June 2015 to develop a discrete choice model of actual site choice (RP model). Other elicited information included characteristics of the current trip, typical characteristics of trip taken to the other types of nature-based recreation, and socioeconomic data.

A total of 572 out of 720 intercepted respondents completed the survey (79% response rate). The sample used in the study was however reduced to 405 day-trippers who resided within 250 kilometers from the furthest waterfall site included in the study and traveled to the intercept site by personal vehicle. Nevertheless, eight respondents were dropped due to possible protest bias detected by the screening question. This resulted in the final sample of 397 and a total of 1,588 choice scenarios used for estimation of SP model. For a comparison of two data sets generated from internet and in-person interview surveys, a random sample of 128 respondents who provided answers for 512 choice scenarios were drawn from the final sample and used for estimations of models which utilized this data (SP and joint SP-RP models). This strategy was employed to ensure the closest possible comparability with the internet survey sample when controlling for scale differences in the joint SP-RP models.

As for the RP sample, a total of 938 day-trips were actually undertaken by 405 identical samples to the set of 10 waterfall sites defined as the choice set. Similar to Haener, Boxall, & Adamowicz (2001), holdout samples for comparison of the trip distributions in prediction tests were generated by randomly drawing 100 respondents from the final sample. The remaining 305 respondents were then used for estimations of models which utilized this data (RP and joint SP-RP models).

The site characteristic data used for estimation of models which utilized RP data was based mainly on the Carrying Capacity (CC) study reports of DNP (2006). Unofficial CC information of Ched Sao Noi falls was provided by its staff through personal communication. Personal site survey was conducted to collect the missing site characteristic data of Nang Rong falls, Muak Lek falls, and E-To falls. The waterfall sites included in the choice set for RP model were: Takro falls (TKO), Ched Sao Noi falls (CSN), Heaw Narok falls (HNR), Heaw Suwat falls (HSW), Nang Rong falls (NRG), Sa Rika falls (SRK), Pha Kluay Mai falls (PKM), Krong Kaew falls (KK), Muak Lek falls (MLK), and E-To falls (ETO). Thus, it was assumed that the sampled recreationist chose to visit the waterfall from a fixed choice set of 10 waterfalls on a given trip occasion when estimating models which utilized RP data.

The sample for internet survey mode was recruited through the popular social media during October, 2015 due to the lack of sample source for locating potential members of waterfall recreationists.[4] The link to a designated webpage of the online survey service provider was posted on the first author's Facebook® page. The privacy status of this particular Facebook® post was set as public to allow participants to share this particular post to their Facebook® friends. The content and appearance of choice experiment questions

[4] As previously discussed, confounding sample composition effects seem to be one of the most difficulties in survey mode study in SP literature. Only Boyle et al. (2016) that strictly control the sample frame to investigate its effects on welfare estimates for implementation of a mail survey. None of survey mode study has been conducted to investigate the effect of different sample frames on other survey modes.

presented to respondents were exactly the same as those of in-person interview. Nevertheless, since the use of mobile phone in doing the survey was highly possible, the whole questionnaire was then shorter than that of in-person interview. The auxiliary questions included only screening questions to identify potential waterfall recreationist, some characteristics of last trip taken to the waterfall, and some general characteristics when taking the trips to natural recreation resources (e.g. beach, mountain, island, etc.)

Since the sample was recruited through participants' interpersonal relations and their connections (called "chain referral sampling" in survey methodology literature), the estimated coefficients would potentially be biased (see Heckathorn (1977); and Heckathorn (2011) for discussion on snowball and other chain referral sampling methods). On the other hand, Hensher, Rose, and Greene (2005) argue that, for SP samples, the scenarios are assigned in a pre-arranged manner suggesting that the analyst has no means to force a decision maker to select a specific alternative. The SP data itself is therefore sampled randomly. Note that this matter is not yet fully resolved in the literature, and, perhaps, is the promising subject for future research.

A total of 202 respondents participated in the online survey. Sampling frame was defined as the respondents who took at least a trip to the waterfall site in the last five years. The respondents classified as non-potential waterfall recreationist were then excluded from the analysis. This left the final sample of 151 respondents who provided answers for about 604 choice scenarios. However, since the number of respondents answering each block was unequal, the sample of 16 respondents was randomly drawn from each block of the final sample to be used as the estimation sample. The process of equally drawing identical number of samples from each block was conducted to minimize any possible choice sets bias. Therefore, only 128 respondents who provided answers for 512 choice scenarios were used for estimations of models which utilized this data (SP and joint SP-RP models).

Site Choice Model: Theory and Econometric Specifications

Recreationists' site choices are analyzed using the discrete choice conditional logit model which is based on the random utility maximization (RUM) framework (McFadden, 1973; Hanemann, 1984). Under RUM framework, recreationist chooses to visit the waterfall site that provides him/her the highest utility. The utility function is assumed to be deterministic for recreationist (V) but contains some unobservable components (ε) which are treated as random variables by the researcher.

$$U = V + \varepsilon \qquad (1)$$

where V is deterministic indirect utility function observed by the researcher, and ε is stochastic components unobservable to the researcher. Assume that the utility that a recreationist derives from visiting the waterfall site is associated with the attributes of the waterfall. The indirect utility of recreationist n to waterfall site j can then be characterized by the following additively-separable linear-in-attributes form

$$V_{nj} = \beta_k \boldsymbol{X}_{nj} \qquad (2)$$

where X is a vector of k attributes associated with waterfall site j and β is a preference coefficient vector. Specifically, in the case of SP model, the utility that recreationist derives from choosing alternative from a fixed choice set of a pair of designed hypothetical waterfall sites and an option of stay home in the choice experiment is assumed to be explicitly expressed as

$$\begin{aligned} U_{A\ falls} = {} & \beta_1 DIST_A + \beta_2 FEE_A + \beta_3 TYPE_A + \beta_4 FLOW_A + \beta_5 SWIMH_A + \beta_6 SWIMM_A \\ & + \beta_7 SWIML_A + \beta_8 PICNH_A + \beta_9 PICNM_A + \beta_{10} PICNL_A + \beta_{11} ENTRY_A \\ & + \beta_{12} NATQ_A + \beta_{13} WW_A + \beta_{14} TRAIL_A + \beta_{15} SIGN_A + \beta_{16} HWAY_A + \varepsilon_A \end{aligned} \quad (3)$$

$$\begin{aligned} U_{B\ falls} = {} & \beta_1 DIST_B + \beta_2 FEE_B + \beta_3 TYPE_B + \beta_4 FLOW_B + \beta_5 SWIMH_B + \beta_6 SWIMM_B \\ & + \beta_7 SWIML_B + \beta_8 PICNH_B + \beta_9 PICNM_B + \beta_{10} PICNL_B + \beta_{11} ENTRY_B \\ & + \beta_{12} NATQ_B + \beta_{13} WW_B + \beta_{14} TRAIL_B + \beta_{15} SIGN_B + \beta_{16} HWAY_B + \varepsilon_B \end{aligned} \quad (4)$$

$$U_{home} = \alpha_{home} + \varepsilon_{home} \quad (5)$$

where α_{home} is an alternative specific constant which captures the preference for staying at home. Theoretically, if the distribution of stochastic component ε is assumed to be independently and identically *Type-I* extreme value over time, recreationists, and alternatives, McFadden (1973) shows that the conditional choice probability of visiting site i (that is, in our case, either A falls, B falls, or Stay at home) among alternative site $j=1,2,3,\ldots,J$ of the sampled recreationist n for the Conditional Logit model can be expressed as[5]

$$Prob_n(i) = \frac{e^{\mu v_{ni}}}{\sum_{j=1}^{J} e^{\mu v_{nj}}} \quad (6)$$

where μ is a scale parameter and J is the total number of choice set C. When the model is estimated using a single set of data, μ cannot be identified and, therefore, is confounded with the preference coefficient vector (Swait & Louviere, 1993). Typically, when a single set of data (either SP or RP) is used to estimate a model, μ is assumed to be unity because it has no effect on the utility levels (Adamowicz et al., 1994).

Standard maximum likelihood (ML) estimation typically provided in the most standard econometric packages is used to estimate the preference coefficients, β_k. Given the sample size N, the following log-likelihood function (LL) is maximized

$$LL(\beta) = \sum_{n=1}^{N} \sum_{j \in \mathbf{C}} f_{nj} \ln\left(Prob\{j \mid \beta\}\right) \quad (7)$$

---

[5] The conditional logit model implicitly assumes that: (1) all respondents have the same preference structure, (2) choices conform to the Independence from Irrelevant Alternatives (IIA) assumption, and (3) stochastic are independent over time. These assumptions are fairly restrictive and can be altered by using other less restrictive models such as multinomial probit model or mixed logit model. However, since RP data used in this study was from the intercept survey that not all sites in the choice set were sampled, none of multinomial probit or mixed logit models, to the best of our knowledge, is capable of correcting this sort of choice-based sampling bias.

where f_{nj} is the frequency of choice *j* chosen by recreationist indexed by *n=1,…,N* from the choice set **C** = {*A falls, B falls, stay at home*} and zero otherwise (Ben-Akiva and Lerman, 1985; Haab and McConnell, 2002); and $Prob\{ j \mid \beta \}$ is the probability of recreationist *n* choosing alternative *j*.

For estimation of RP model, the utility that the sampled recreationist derives from taking a trip to a site in one of the fixed choice set of 10 waterfalls is assumed to be expressed as

$$U_j = \beta_1 DIST_j + \beta_2 FEE_j + \beta_3 TYPE_j + \beta_4 SWIM_j + \beta_5 PICN_j + \beta_6 ENTRY_j + \beta_7 NATQ_j + \beta_8 WW_j + \beta_9 HWAY_j + \varepsilon_j$$
$$; j = TKO, CSN, HNR, HSW, NRG, SRK, PKM, KK, MLK, ETO \quad (8)$$

The shortest travel distance is calculated from the proxy place of respondent's residence to the study sites using the Google Map® website. Other attribute levels are coded to match with actual characteristics of waterfall sites. In order to construct a fair comparison between RP and SP data, as many attributes as possible used for estimation of SP model are included in the RP model. Except swim area and picnic spots attributes, the number of levels of categorical variables used in estimation of RP and SP models are exactly the same. Because the total size of picnic area data is not available for all alternative sites in the choice set and the use of data of similar waterfall site in the region is difficult, the number of picnic spots variable is coded as 2-level (available vs non-available) attribute. For swim area attribute, because the attribute is severely collinear with the other attributes and most of the coefficients cannot be estimated when levels of swim area variables (*SWIM_H, SWIM_M, SWIM_L*) are included, the swim area attribute is then re-coded to 2-level *SWIM* variable (available vs non-available).[6,7] The natural trail attribute is also not included due to its high colinearity with other attributes. For water flow and interpretative media attributes, they are not included due to the lack of information.

Because the intercept sample has a higher level of avidity than the waterfall recreationists randomly chosen from the general population, the estimation of probabilistic choice models from choice-based sampling data yields biased parameter estimates (Manski and Lerman, 1977). To correct for the choice-based sampling bias, McFadden's Intercept & Follow (I&F) estimator is used (McFadden, 1996), as it matches with the sampling strategy used in field survey. More detail of this estimator can be found in

[6] Variance inflation factors (VIF) were calculated to investigate multi-colinearity problem.

[7] Although the site-specific total swim area (m^2) data were also not available for all alternative sites in the choice set, the size of swim area was also tried. As opposed to the case of picnic spots attribute, the data of similar waterfall site in the region were possibly referred to and used as a proxy for the true data. Nevertheless, because the available swim area data did not match with the actual swim area observed during the survey, the data was used as indicative measure in categorizing the level of swim area attribute only.

Kamolthip (2016). The log-likelihood function of McFadden's I&F estimator of the pooled sample can be expressed as

$$LL_{I\&F}(\beta)=\sum_{n=1}^{N}\sum_{j\in C}f_{nj}\ln\left(Prob\{j\mid\beta\}\right)+\sum_{n=1}^{N}\sum_{r=1}^{R}\frac{\delta_{nr}}{w_r}ln\left(\frac{Prob\{r\mid\beta\}}{\sum_{r=1}^{R}Prob\{r\mid\beta\}}\right) \tag{9}$$

where f_{nj} is the frequency of trips taken by recreationist indexed by $n=1,\ldots,N$ to site j among the choice set $C=\{TKO, CSN, HNR, HSW, NRG, SRK, PKM, KK, MLK, ETO\}$ during the defined period other than the intercept trip; $Prob\{\ j\mid\beta\ \}$ is the probability of a recreationist n choosing waterfall site j; W_r is sampling correction weights for intercept site r and equal to the ratio of the sample probability for intercept site r, indexed by $r=1,\ldots,R,$ to the waterfall recreationist population probability of visiting site r; $Prob\{\ r\mid\beta\ \}$ is the probability of a recreationist n choosing the intercepted waterfall site r; and δ_{nr} = 1 when recreationist n is intercepted at site r and zero otherwise.

As mentioned previously, the scale parameter μ is confounded with the preference coefficient vector and unavoidably assumed to be unity when estimating a single set of data (either RP or SP). Nevertheless, the ratio of scale parameters can be determined when the two sets of data are jointly estimated since both RP and SP models identically base the process of recreational site choice on the site characteristics (Adamowicz et. al., 1994; Louviere et al., 2000). The joint SP-RP model is then introduced as one of alternative models to investigate the effect of internet SP data taking into account the scale parameters of both SP and RP data. When estimating the joint SP-RP model, the utility that recreationist derives from choosing alternative from a fixed choice set of each model is assumed to be expressed as

$$\begin{aligned}U_j=\alpha_{home}&+\beta_1DIST_j+\beta_2FEE_j+\beta_3TYPE_j+\beta_4FLOW_j+\beta_5SWIMH_j\\&+\beta_6SWIMM_j+\beta_7SWIML_j+\beta_8PICNH_j+\beta_9PICNM_j+\beta_{10}PICNL_j\\&+\beta_{11}ENTRY+\beta_{12}NATQ_j+\beta_{13}WW_j+\beta_{14}TRAIL_j+\beta_{15}SIGN_j+\beta_{16}HWAY_j+\varepsilon_j\\&\qquad;j^{RP}=TKO,\ CSN,\ HNR,\ HSW,\ NRG,\ SRK,\ PKM,\ KK,\ MLK,\ ETO\\&\qquad;j^{SP}=HOME,\ A\ FALLS,\ B\ FALLS\end{aligned} \tag{10}$$

Note in this case all attributes equal to zero when j = *HOME* as explicitly presented by Equation (5). Since the addition of SP data in the joint SP-RP model reduces the colinearity in the RP data (Adamowicz et al, 1994), it is then possible to use the 4-level swim area (*SWIM_H, SWIM_M, SWIM_L*) and the *TRAIL* attributes for the RP data. Thus, the estimation includes 11 attributes common across the RP and SP data plus 5 SP-specific attributes (*FLOW, PICNH, PICNM*, *PICNL,* and *SIGN*) and one SP's ASC for *HOME*.

When estimating the joint SP-RP models, the log-likelihood function of joint estimation suggested by Louviere et al. (2000) is adjusted by adding the second term in McFadden's I&F estimator Equation (9) in order to correct the sampling bias arisen from the choice-based sample. Following log-likelihood function is used for estimation of joint SP-RP models reported in this study

$$LL_{I\&F}\left(\beta, Z^{RP}, Z^{SP}, \tau\right) = \sum_{n=1}^{N^{RP}} \sum_{j \in C^{RP}} f_{nj}^{RP} \ln\left(Prob\left\{j \mid \beta, Z^{RP}\right\}\right) + \sum_{n=1}^{N^{RP}} \sum_{r=1}^{R} \frac{\delta_{nr}}{W_r} \ln\left(\frac{Prob\left\{r \mid \beta, Z^{RP}\right\}}{\sum_{r=1}^{R} Prob\left\{r \mid \beta, Z^{RP}\right\}}\right) + \sum_{n=1}^{N^{SP}} \sum_{j \in C^{SP}} f_{nj}^{SP} \ln\left(Prob\left\{j \mid \beta, Z^{SP}, \tau\right\}\right) \quad (11)$$

where f_{nj}^{RP} and f_{nj}^{SP} are the frequency of trip taken by recreationist indexed by $n=1,\ldots,N$ to site j from the choice set among the choice set C^{RP}={*TKO, CSN, HNR, HSW, NRG, SRK, PKM, KK, MLK, ETO*} and C^{SP}={*HOME, A FALLS, B FALLS*}, respectively; $Prob\{ j \mid \beta, Z^{RP}\}$ and $Prob\{ j \mid \beta, Z^{SP}, \tau \}$ are the probabilities of recreationist n choosing alternative j in the RP and SP samples, respectively; W_r is sampling correction weights for intercept site r and equal to the ratio of the sample probability for intercept site r, indexed by $r=1,\ldots,R,$ to the waterfall recreationist population probability of visiting site r; $Prob\{ r \mid \beta \}$ is the probability of a recreationist n choosing the intercepted waterfall site r; δ_{nr} = 1 when recreationist n is intercepted at site r and zero otherwise; β is coefficient vector common between SP and RP data which is restricted to be equal in the estimation (Haener et al., 2001); Z^{RP} and Z^{SP} are coefficient vector associated with attributes unique to the RP and SP samples; and τ is relative scale parameter μ_{SP} / μ_{RP} of which the scale of RP data is normalized to one.

Preliminary Estimation

Table 2 reports the results of three separate estimations of site choice models when the whole final sample of each data set (RP data, in-person interview SP data, and internet SP data) is fully utilized. Since the subsequent analyses use lesser number of choice observations in comparing underlying preference structure and predictive ability, the results presented in this table are more preferable for the interpretation of coefficient estimates. In this study, statistical package NLOGIT 5 is used for all estimations.

Categorical variables were coded using effect codes (Louviere, 1988). For effect coding, each column is assigned 1 for the level represented and all columns are assigned -1 for the base level. The effect codes then result in one fewer coefficient than the number of levels. The interpretation of their coefficients is that each level takes the utility associated with the coefficient and the base level takes the utility associated with the negative sum of the coefficients of the other remaining levels (Adamowicz et al., 1994; Haener et al., 2001). Furthermore, to conduct the closest possible comparison, alternative specific constants (ASCs) for waterfall sites were not included in any models that used RP data because they could not have been estimated by the SP data (see Haener et al., 2001).[8]

---

[8] In contrast, Louviere et al. (2000) suggest that the RP ASCs should be included in the models that will subsequently be used for the prediction. This matter is however not yet fully resolved.

For RP model, all attributes except for the quality of walkway are significant. As expected the coefficients for travelling distance and entrance fee are significant and negative, supporting the underlying idea of travel cost framework. The sign of each coefficient indicates that the sampled recreationists have positive preferences for the waterfall which has obvious drops and there are picnic spots available for recreationists within the site. The positive coefficients of “Natural park” and “Highway” suggest that the sampled recreationists prefer the waterfall site which is located in the national park and convenient for travelling to the site. The coefficient is negative for “Entry distance”, so it is more likely that the sampled recreationists dislike the waterfalls that they have to walk for a long distance from entrance to the waterfall body. The negative coefficient for availability of swim area is however not expected and contrary to the hypothesis in literature.[9]

Table 2: Coefficient estimates from final sample of each data set

	RP		SP-F2F		SP-IN	
Attribute	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
SP ASC for Home			-3.19508***	0.19111	-2.10805***	0.21816
Travel Distance	-0.02458***	0.00205	-0.00062	0.00063	-0.00094	0.00104
Entrance Fee	-0.00641**	0.00319	-0.00295**	0.0012	-0.00536***	0.00201
Type of waterfall	0.16340***	0.05285	0.02016	0.02701	0.06801	0.0455
Water flow			-0.0078	0.02708	0.02207	0.04576
Entry Distance	-0.29805***	0.081	-0.10002***	0.02729	0.02403	0.04611
National Park	0.33204***	0.0913	0.10983***	0.02684	0.17349***	0.04621
Highway	0.37638***	0.06779	-0.02677	0.02703	0.04415	0.04656
Plenty of Swim Area [a]	-1.47589***	0.36234	0.27899***	0.05814	0.13282	0.09549
Moderate Swim Area			0.23046***	0.05467	0.33776***	0.09283
Few of Swim Area			-0.0239	0.06044	-0.08021	0.10042
Quality of Walkway	-0.05425	0.07685	-0.05946**	0.02725	-0.05518	0.04616
Plenty of Picnic Spots [a]	1.64534***	0.36271	0.09894*	0.05624	0.13856	0.09168
Moderate Picnic Spots			0.02345	0.05817	0.01414	0.09896
Few of Picnic Spots			0.04049	0.05204	0.15321*	0.08746
Natural Trail			0.04688*	0.027	0.04114	0.0454
Interpretative Media			0.08244***	0.02703	0.14450***	0.04569
Choice observations	938		1588		604	
Log-likelihood	-1483.58918		-1178.29738		-510.12676	

Notes: 1. ***, **, * indicate significance at 1%, 5%, 10% level, respectively.

2. [a] indicates that, for RP model, this coefficient represents “availability” of associated attribute, instead of the levels.

[9] This result is probably caused by the type of variable used in modelling. When the size of area (m^2) is used instead of dummy variable for swim area, the coefficient is positive and significant as expected. However, because the available size of swim area data is doubtful, we decide to use the RP model as reported in the Table 2.

In the in-person interview SP model (hereinafter, SP-F2F), although the coefficients of both entrance fee and travelling distance are negative as expected; only the coefficient of entrance fee is significant for travel cost-related variables. For significant attributes that are common with RP model, except for levels of swim area, the coefficient vector shows that the pattern of preferences across the attributes is similar to that observed in the RP model. The positive coefficients of “Natural trail” and “Interpretative media” suggest that the recreationists in the in-person interview SP data prefer these attributes. The coefficient is negative for the quality of walkway attribute, so the recreationists in this data source more likely dislike the developed walkway. As opposed to RP data, the coefficients are positive for levels of swim area. This result is however consistent with the result of RP model when the size of area is used, as mentioned previously (see footnote 10). The other significant coefficients have the same as those observed in the RP model, and thus have the same interpretations.

For the internet SP model (hereinafter, SP-IN), the number of significant coefficients is less than that observed in SP-F2F model. As the reader will see in the next section, this could be explained by the less number of choice observations used in the estimation. Since all of significant coefficients have the same sign as those observed in the SP-F2F model, but the weights on the attributes are quite different, the same interpretation is thus applied for each attribute.

Validity Tests

Preference Homogeneity Test

To test whether the two SP data sets (internet SP and in-person interview SP) yield consistent information on the underlying preferences of recreationists, five different models were estimated. Specifically, three estimations of each single set of RP, internet SP, and in-person interview SP data and two estimations of pooled RP-internet SP data (hereinafter, J-SPIN) and pooled RP-person interview SP data (hereinafter, J-SPF2F) were estimated. The resulting coefficient estimates are reported in Table 3.

In testing the hypothesis, Swait & Louviere (1993) suggest the following steps: (1) Estimate separate conditional logit models for each data set to obtain log likelihood (in our case, L^{RP} and L^{SP}); (2) Estimate a conditional logit mode from the pooled data to obtain L^{Joint}; and (3) Calculate the chi-square statistic for the hypothesis from $-2((L^{RP}+L^{SP}) - L^{Joint})$ and compare with the critical value for the α =0.05 significance with $K^{RP} + K^{SP} - K^{Joint}$ degrees of freedom (see Louviere et al., 2000, for more detail and discussion).[10]

For RP data, the estimation sample of 305 respondents was randomly selected from the final sample of RP data to minimize possible over-weighting of RP observations in the joint model. The remaining 100 respondents were set aside as the holdout sample to test against the trip distribution predicted by the results

[10] To restrict the chi-square to be positive $-2(L^{Joint} - (L^{RP}+L^{SP}))$ can be used interchangeably. The original paper of Swait & Louviere (1993) applies this form. The test statistic presented in this paper is based on Louviere et al. (2000).

obtained from the estimations of five different models, as an external validity test. In doing this, a set of 100 respondents (~25% of the final sample) was randomly drawn from the final sample of RP data such that the share of sampled recreationists for each intercept site in the holdout sample was approximately proportional to the actual recreationist population shares (this is called pseudo-random design in the choice-based sampling literature (McFadden, 1996)).[11] This was conducted to minimize the effect of choice-based sampling and allow the direct comparison of the distributions of predicted and actual trips. The remaining sample of 305 respondents was then used for estimation of models which utilized RP data (RP, J-SPIN, J-SPF2F models).

For internet SP data, there was a problem of unequal number of respondents answering each block of choice experiment for internet SP data. To minimize over-weighting of some particular blocks, the sample of 16 respondents per block was randomly drawn from the final sample of internet SP data and used for estimation. Therefore, a total of 128 respondents who provided answers for 512 choice scenarios were used for models which utilized internet SP data (SP-IN and J-SPIN models). Despite the higher number of respondents per block of choice experiment for in-person interview SP data in hand; the randomly selected samples were used for estimations of models to ensure a fair comparison. An exactly identical process used for internet SP data was applied to in-person interview SP data. The samples of 16 respondents per block were randomly drawn from the final sample of in-person interview SP data. A total of 128 respondents who provided answers for 512 choice scenarios were then used for models which utilized in-person interview SP data (SP-F2F and J-SPF2F models).

The process of randomly drawing the sample from RP data, internet SP data, and in-person interview SP data was repeated for forty different sets of holdout (RP only) and estimation (RP, internet SP, in-person interview SP) samples. All of the RP, SP-IN, and SP-F2F models were re-estimated for each new draws of the estimation sample from the final sample of their associated data sets. For J-SPIN and J-SPF2F models, each draw of the estimation sample from relevant data sets were pooled and used for estimation of models. In total, forty sets of coefficient vectors and statistics for each model were derived for analysis.

Because each new draw resulted in new composition of both estimation sample and holdout sample, the total number of trips taken by the recreationists in each set of the sample then changed from replication to replication. A representative set of coefficient estimates selected from forty replications was then reported to simplify the presentation. For the set of models reported in Table 3, the models which utilized RP data (RP, J-SPF2F, J-SPIN) contained information from 305 recreationists who took 706 trips. The remaining 100 recreationists included in the holdout sample took 232 trips.[12]

---

[11] Since not all alternative sites in the choice set were surveyed, in this case the actual recreationist population share for each intercept site was normalized by the sum of their shares. In other words, the same sampling correction weight was applied.

[12] The total numbers of trips taken in forty holdout samples were from 200 to 281 trips. On average, 100 recreationists included in each holdout sample took 232 trips.

Table 3: Coefficient estimates for five recreation site choice models using representative set of randomly drawn estimation sample

	RP	SP-F2F	SP-IN	J-SPF2F [c]	J-SPIN [c]
SP ASC for Home	–	-3.31806***	-2.03436***	-3.27223***	-1.97566***
		(0.33657)	(0.23332)	(0.30188)	(0.18381)
Travel Distance	-0.02389***	-0.00173	-0.00059	-0.00079**	-0.00059**
	(0.00234)	(0.00111)	(0.00112)	(0.00034)	(0.00028)
Entrance Fee	-0.00748**	-0.00339	-0.00606***	-0.00462***	-0.00404***
	(0.00370)	(0.00214)	(0.00219)	(0.0013)	(0.00145)
Type of waterfall	0.15670**	0.06238	0.04424	0.04853***	0.04201***
	(0.06142)	(0.04839)	(0.04955)	(0.01378)	(0.01509)
Water flow	–	0.10652**	0.01452	0.09898**	0.01595
		(0.04843)	(0.04974)	(0.04642)	(0.04814)
Entry Distance	-0.34730***	-0.13686***	0.02575	-0.02888*	-0.01974*
	(0.09100)	(0.04920)	(0.04979)	(0.01505)	(0.01136)
National Park	0.33302***	0.17180***	0.19400***	0.02956**	0.02973**
	(0.10545)	(0.04827)	(0.04992)	(0.0143)	(0.01348)
Highway	0.38485***	-0.06207	0.04942	-0.05318***	-0.04741***
	(0.07586)	(0.04869)	(0.05023)	(0.01685)	(0.01781)
Plenty of Swim Area [a]	-1.71661***	0.20324**	0.05300	-0.11026***	-0.13713***
	(0.50778)	(0.10173)	(0.10573)	(0.03295)	(0.03434)
Moderate Swim Area	–	0.29652***	0.39228***	0.31529***	0.38864***
		(0.09837)	(0.10053)	(0.09475)	(0.09898)
Few of Swim Area	–	-0.07069	-0.09407	-0.12067***	-0.13760***
		(0.10778)	(0.1084)	(0.036)	(0.03502)
Quality of Walkway	-0.05842	-0.01516	-0.03254	0.01920*	0.01262
	(0.08818)	(0.04844)	(0.04991)	(0.01165)	(0.00865)
Plenty of Picnic Spots [a]	1.86684***	0.06972	0.16336	0.08262	0.12977
	(0.50852)	(0.09954)	(0.10046)	(0.09377)	(0.09799)
Moderate Picnic Spots	–	0.05327	0.01105	0.02817	0.01849
		(0.10522)	(0.10765)	(0.10115)	(0.10407)
Few of Picnic Spots	–	0.10412	0.14017	0.11178	0.13854
		(0.09397)	(0.09408)	(0.09227)	(0.0932)
Natural Trail	–	0.06366	0.03806	0.08932***	0.07483***
		(0.04799)	(0.04949)	(0.02576)	(0.0271)
Interpretative Media	–	0.05675	0.15551***	0.05495	0.16031***
		(0.04832)	(0.04971)	(0.04636)	(0.04811)
Relative Inclusive Value Parameters					
θ^{RP}	–	–	–	0.03417**	0.02546**
				(0.01465)	(0.01192)
θ^{SP}	–	–	–	1 (Fixed) [b]	1 (Fixed) [b]
Choice observations	704	512	512	1216	1216
Log-likelihood	-1107.80049	-372.25958	-433.37895	-1476.99879	-1534.35983

Notes: 1. ***, **, * indicate significance at 1%, 5%, 10% level, respectively.

2. Standard errors are in parentheses.

3. [a] indicates that, for RP model, this coefficient represents “availability” of associated attribute, instead of the levels.

4. [b] indicates that inclusive value parameters for SP data set are normalized to be unity.

5. [c] indicates that all coefficients, except SP ASC for HOME, are confounded with relative inclusive value of RP data.

A Full Information Maximum Likelihood (FIML) method was used for estimations of J-SPF2F and J-SPIN models.[13] Since the scale parameter is assumed to be identical for alternatives in each single set of the data, but different between data set, the structure of pooled SP-RP data is similar to the nests in a Nested Logit (NL) model (Louviere et al., 2000). Such *artificial tree structure* (Hensher & Bradley, 1993) then allows us to estimate an NL model from the two data sources so as to obtain coefficients and relative scale parameter simultaneously.

The significance and sign of all coefficients except "Few of Picnic Spots" of the SP-IN models which utilizes subsample are unchanged when compared to that observed in the same model but utilizes the whole final sample. However, the significance and sign of the coefficients of SP-F2F model which utilizes subsample are quite different when compared to that observed in the same model but utilizes the whole final sample. Furthermore, the magnitudes of the coefficients of SP-F2F model also considerably change. This is probably caused by the large difference between the choice observations used in both estimations, as the subsequent results show that both RP and in-person interview SP data more likely share the same underlying preference structure. More attributes are significant when the RP and SP data are jointly estimated relative to two SP models.

Following the steps mentioned earlier, the associated chi-squared statistic for J-SPF2F and J-SPIN models are 6.12255, 13.6392, respectively. When compared with the critical value at 0.05 significance level with 9 degrees of freedom (χ_c^2 = 16.91898), the hypothesis of the preference homogeneity between RP data and each single set of SP data cannot be rejected. Thus, for this representative set of models, the results suggest that the underlying preference structures of internet SP and in-person interview SP data sets are not statistically different from that observed in the more preferred RP data, when controlling for scale differences.

The preference homogeneity test was then repeated for the remaining 39 sets of estimation sample in order to examine the robustness of these results. The J-SPIN model was successfully estimated in 39 out of 40 replications, including the representative set. For the failed replication, a lack of trip to the Krong Kaew falls in the randomly selected estimation sample of RP data affected the variation in the attribute levels and, consequently, affected the estimation of models that utilized this set of RP data (RP, J-SPIN, and J-SPF2F models). The chi-squared statistics for the J-SPIN models in all of the successful 39 replications suggested that the preference homogeneity hypothesis between RP and internet SP data in each replication could not be rejected. Similarly, the 39 out of 40 replications were successfully estimated for the J-SPF2F model, including

---

[13] When estimating the joint SP-RP model, two estimation methods are available: the Full Information Maximum Likelihood (FIML) and the manual "grid search" methods. The latter one, originally proposed by Swait & Louviere (1993), needs a special code written to estimate the coefficients and the relative scale parameter by manual search. Although their procedure yields consistent estimates, it is however argued that the coefficients and the relative scale parameter derived from the manual method are not efficient, leading to inflated *t*-statistics (Louviere et al., 2000). NLOGIT codes for the FIML and the manual "grid search" methods are available from the author upon request.

the representative set. Nevertheless, only 35 out of 39 replications that we could not reject the hypothesis of preference homogeneity between RP and in-person interview SP data. The chi-squared statistics for the J-SPF2F model in each of the other four replications suggested that the hypothesis of the same underlying preference structure of the sampled recreationists of in-person interview SP data and RP data could be rejected.

The results are quite surprising in that the SP data collected by internet survey have more replications that have the same underlying preference structure as the RP data than the SP data collected by in-person interview. Since the RP and in-person interview SP data were collected in the same time from the same sample frame, it was expected that the in-person interview SP data would undoubtedly have the same underlying preference structure as the RP data. This expectation was also based on a conventional belief that the SP survey which administered in-person would provide superior data quality (Arrow et al., 1993).

Predictive Ability Tests

This section examines the predictive ability of in-person interview and internet data sets. In doing this, the predicted trip distributions of all five models using associated randomly selected estimation sample were compared with the actual trip distribution observed in the holdout sample. However, since the error variances of internet SP and in-person interview SP data were usually larger than the RP error variance, the trip distributions predicted by J-SPIN and J-SPF2F models were also included in the comparisons to present the predictive ability of two data sets when scale (or, equivalently, variance) differences were accounted for.

Horowitz & Louviere (1993) and Haener et al. (2001) provide list of both aggregate level and individual level tests of predictive ability. Based on the literature, four tests and measures were chosen and used to examine the differences in predictive ability of both data sets when compared with the predictive ability of RP data. The first two tests were conducted at aggregate level: (1) the sum of absolute errors (SAE); and (2) the degree of correlation between the predicted and observed aggregate trip distributions (r_a). The next three tests assessed predictive ability at individual level. These included: (3) the degree of correlation between the predicted and observed trips pooled across all recreationists (r_i); and (4) the mean of individual-specific correlation (r_m).

Table 4 reports the aggregate predicted trip distributions of the five models using the representative set of randomly drawn estimation sample and the aggregate observed trip distribution of corresponding RP holdout sample. In calculating the predicted trip distribution of each model, the probability vector of each recreationist visiting ten alternative waterfall sites was first calculated based on the estimated coefficient vectors, site-specific data for the attributes, and recreationists' estimated travel costs. Predicted trip distribution was then derived by multiplying individual probability vector with their total number of trips. These predicted trip distributions were subsequently used to calculate individual level statistics (r_i, r_m). The aggregate predicted trip distributions were calculated by summing the individually predicted trips to each alternative waterfall site over the recreationists in the estimation sample. The resulting aggregate trip distributions were subsequently used to calculate aggregate level statistics (SAE, r_a).

Table 4: Aggregate actual vs predicted aggregate trip distributions

	Actual Trips [a]	Predicted Trips [a]				
Waterfall Sites	Holdout Sample	RP	SP-F2F	SP-IN	J-SPF2F	J-SPIN
Takro	24	14.3192	35.3711	25.0612	20.3276	20.1378
Ched Sao Noi	56	48.8812	34.6986	33.4585	53.2572	53.3099
Heaw Narok	26	13.5049	14.0314	19.9813	18.2777	18.2720
Heaw Suwat	22	19.7559	18.4927	20.0637	21.0472	20.9001
Nang Rong	38	37.7237	25.3014	25.4819	35.968	36.1736
Sa Rika	29	48.7726	26.9664	24.7192	45.2369	44.8830
Pha Kluay Mai	4	10.2843	14.0906	21.1370	4.03777	4.6105
Krong Kaew	1	0.5265	23.5330	23.8084	0.47678	0.3416
Muak Lek	28	24.2958	19.6057	20.7528	28.9398	29.0819
E-To	4	13.9358	19.9091	17.5360	4.43113	4.2896
Predictive Ability Measures						
SAE	–	71.9855	119.808	109.085	35.2911	35.7299
r_a	–	0.83276	0.54973	0.73155	0.93677	0.93833
r_i	–	0.51192	0.34436	0.35094	0.53397	0.53464
r_m	–	0.34597	0.21157	0.25247	0.38206	0.38255

Note: 1. [a] indicates that the total trips = 232

The trip distribution predicted by the RP model appears to be more accurate than SP-F2F and SP-IN models. Note that the attribute levels of "*Water Flow*", "*Interpretative Media*", and "*Picnic Spot*" were unavoidably set to equal to zero due to the lack of site-specific information. This setting terribly affected the derived probability vectors of SP-F2F and SP-IN models, and, therefore, resulted in poor prediction performances. Nevertheless, when scale differences are accounted for, the J-SPF2F and J-SPIN models seem to be more accurate than the RP model. The more systematic measures at aggregate level (SAE and r_a) are used to compare these distributions.

The trip distribution predicted by the RP model appears to be more accurate than both SP-F2F and SP-IN models. Note that the attribute levels of "*Water Flow*", "*Interpretative Media*", and "*Picnic Spot*" were unavoidably set to equal to zero due to the lack of site-specific information. This setting terribly affected the derived probability vectors of SP-F2F and SP-IN models, and, therefore, resulted in poor prediction performances. Nevertheless, when scale differences are accounted for, the J-SPF2F and J-SPIN models seem to be more accurate than the RP model. The more systematic measures at aggregate level (SAE and r_a) are used to compare these distributions.

The sum of absolute error which gives equal weight to all errors can be calculated as

$$SAE = \sum_{i=1}^{J} \left| \left(\hat{N}_i - N_i \right) \right| \tag{12}$$

where $\hat{N}_i$ is the total number of predicted trips to waterfall site *i*, N_i is the total number of observed trips to waterfall site *i*, and *J* is the total number of alternative waterfall site in the choice set.

At aggregate level, the RP model has the lowest error in prediction (SAE) when compared with SP-F2F and SP-IN models (71.9855, 119.808, and 109.085, respectively). The SAE value of J-SPF2F model (35.2911) slightly differs from that of J-SPIN model (35.7299), and both are lower than that of RP model. The aggregate correlation coefficient (r_a) of the RP model (0.83276) is higher than that of SP-F2F and SP-IN models (0.54973 and 0.73155, respectively). Note that although the r_a coefficient of the SP-IN model is substantially higher than that of the SP-F2F model, the predictive performances of both data sets are not different, when the scale differences are accounted for in the J-SPF2F and J-SPIN models (0.93677 and 0.93833, respectively). The findings for this set of models and holdout sample indicate that, at aggregate level, the predictive ability of SP-F2F data is not superior to that observed from the SP-IN data.

More measures were also calculated to further compare these distributions at individual level. The predicted and observed trip vectors of all individuals in the estimation sample were first combined. The overall correlation coefficient (r_i) was then calculated across these combined information. The mean of individual-specific correlation coefficient (r_m) introduced by Haener et al. (2001) was also used in comparing the predictive ability of the five models. Using each individual's predicted and observed trip vectors; this measure determines the mean of individual-specific correlation coefficients. The mean of individual-specific correlation coefficient (r_m) is calculated as

$$r_m = \frac{1}{N}\sum_{n=1}^{N}\frac{\operatorname{cov}(X_{n,P}, X_{n,O})}{\sqrt{\operatorname{var}(X_{n,P})}\sqrt{\operatorname{var}(X_{n,O})}} \tag{13}$$

where *n* is individual indexed by *n=1,…,N*; and $X_{n,P}$ and $X_{n,O}$ are the predicted and observed trip vectors for individual *n*, respectively.

Consistent with the predictive ability tests at aggregate level, the RP model has the higher r_i (0.51192) and r_m (0.34597) values than that of SP-F2F (0.34436 and 0.21157, respectively) and SP-IN (0.35094 and 0.25247, respectively) models. However, when compared with r_i and r_m of J-SPF2F (0.53397 and 0.38206, respectively) and J-SPIN (0.53464 and 0.38255, respectively) models, the RP model is inferior. Thus, at individual level, the SP-F2F data is also not superior to that observed from the SP-IN data.

As the robustness examination, the predictive ability tests were repeated for 35 replications that the hypothesis of preference homogeneity could not be rejected. The mean values of the predictive ability measures across these 35 different sets of holdout and estimation samples are reported in Table 5. As also observed in the representative replication, the RP model has, on average, the lowest SAE when compared with SP-F2F and SP-IN models (92.1768, 145.618, and 127.146, respectively). Although a comparison between two SP data sets indicates that the internet SP data outperform the in-person interview SP data, the average SAEs of the J-SPF2F (61.2489) and J-SPIN (62.1562) models indicate no difference in their performances, when the scale differences are accounted for. Similar results are observed when the aggregate correlation coefficient (r_a) measures are used. The r_a coefficient of SP-IN model (0.67165) is substantially higher than that of SP-F2F model (0.34990), but lower than that of RP model (0.76790). The difference in the predictive performances of two SP data sets disappears when the scale differences are accounted for in J-SPF2F and J-SPIN models (0.86720 and 0.86525, respectively).

Table 5: Summary statistics for the predictive ability measures across forty replications

Test	RP	SP-F2F	SP-IN	J-SPF2F	J-SPIN
			Mean values at Aggregate Level [a]		
SAE	92.1768	145.618	127.146	61.2489	62.1562
	(71.9855,119.193)	(119.808,191.570)	(102.141,162.438)	(33.8745,82.4416)	(35.5253,84.7767)
r_a	0.76790	0.34990	0.67165	0.86720	0.86525
	(0.60317,0.87473)	(0.05467,0.57834)	(0.52657,0.79004)	(0.73611,0.94892)	(0.73405,0.94508)
			Mean Values at Individual Level [a]		
r_i	0.53328	0.31958	0.36726	0.55918	0.55894
	(0.37748,0.73530)	(0.21420,0.46784)	(0.27042,0.50014)	(0.42442,0.75615)	(0.42343,0.75572)
r_m	0.32950	0.13564	0.25563	0.36735	0.36681
	(0.29117,0.37166)	(-0.00087,0.21984)	(0.19181,0.31704)	(0.33344,0.40386)	(0.33249,0.40420)
			Mean of Relative Inclusive Value		
θ_{RP}/θ_{SP}	–	–	–	0.0250 [b]	0.0307 [b]
				(0.0027,0.0493)	(0.0222,0.0411)

Notes: 1. [a] indicates that minimum and maximum values of relevant statistics are respectively shown in parentheses.

2. [b] indicates that inclusive value parameters for SP data set are normalized to be unity.

At individual level, the r_i coefficient measure indicates no difference in the predictive performance between two SP data sets. The r_i coefficients of SP-F2F model (0.31958) slightly differs from that of SP-IN model (0.36726), but both are outperformed by the RP model (0.53328). Lastly, the RP model has the highest r_m coefficients when compared with SP-F2F and SP-IN models (0.32950, 0.13564, and 0.25563, respectively). The r_m coefficients of SP-F2F model is substantially lower than that of SP-IN model. Nevertheless, no difference is observed when the scale differences are accounted for in J-SPF2F and J-SPIN models (0.36735 and 0.36681, respectively).

As expected, the RP data is superior to each single set of SP data in modeling demand for recreation. The findings in this section, however, suggest that the predictive performances of the models that use SP-F2F data are not always superior to that observed from the SP-IN data. Nevertheless, when the scale differences are accounted for, the joint estimations of RP and each single set of SP data perform the best prediction.

Discussion

The results from both preference homogeneity and predictive ability tests consistently suggest that the recreational site choice models that use in-person interview SP data are not superior to the models that use internet SP data. The findings do not support what is often assumed that the SP survey which administers in-person would provide superior data quality.

A close examination on the relative scale parameters sheds some light on this matter. Following Equation (14) (from Louviere et al., 2000, p.242) presents the relationship between the scale parameters (μ), inclusive values (θ), and the variance (σ^2) of two data sources (in our case, RP and SP data):

$$\frac{\sigma^2_{RP}}{\sigma^2_{SP}} = \frac{\pi^2/6\mu^2_{RP}}{\pi^2/6\mu^2_{SP}} = \frac{1/\mu^2_{RP}}{1/\mu^2_{SP}} = \left(\frac{\theta_{RP}}{\theta_{SP}}\right)^2 \qquad (14)$$

The mean of relative inclusive values across all replications of J-SPF2F model is 0.0250, compared with 0.0307 of J-SPIN model. This means that the variance of the RP data is about 0.06% of the variance of the in-person interview SP data and 0.09% of the variance of the internet SP data. The results suggest that the variance of each single set of SP data is considerably larger than that observed in the RP data, and the variance of the in-person interview SP data is about 1.5 times that of the internet SP data.

Swait & Adamowicz (2001a) suggest that the decision environment and choice task characteristics can influence the variance of the choice data. Specifically, they interpret the scale parameter as a representation of the ability to choose and explicitly express the scale parameter as a function of effort (E) and complexity (H) interaction, $\mu(EH)$. The variance of the choice data would thus increase when the respondents' ability to choose declines or the task complexity increases (see, also, Swait & Adamowicz, 2001b).

Given an identical content and appearance of the choice experiment questionnaire presented to the respondents in both survey modes, a possible explanation for the differences in variance between two SP data sets is the administration of the survey. Two factors observed as the differences between our in-person interview SP data and internet SP data are discussed: (1) the in-person vs self-administered setting; and (2) the location setting. First, the presence of interviewers in the in-person setting is generally believed to motivate the respondents in answering the questions. They can also help explain the survey, making the respondents better understand the sophisticated hypothetical questions. These two advantages are believed to reduce the sub-optimal decisions caused by insufficient effort put in making the utility-maximizing choices (also called "*satisficing*" in survey methodology literature). Lindhjem & Navrud (2001b), on the other hand, suggest that the in-person setting probably have an adverse effect on the SP data (e.g. longer time, pressure felt by respondents), thus inducing satisficing. In our case, the findings somehow support the latter argument rather than the former one. As suggested by Lindhjem & Navrud (2001b), our respondent of in-person interview survey might either feel that she was coerced to answer the questions, or feel that the survey would take her time too long, and then shortcut the response process. In contrast, our respondent of internet survey faced none of these problems as she could answer in her own time without any pressure from interviewer.

Second, the in-person interview SP data used in this study was conducted on-site, whereas the internet survey could be taken at the convenient places, depending on the respondents. Lindhjem & Navrud (2001b) suggest that respondent in in-person interview on-site or in other public locations may feel too rushed, resulting in sub-optimal decisions. Haener et al. (2001) present an example of the on-site interview with low interruption. Their findings suggest that the SP survey which was administered in-person during group meetings generate superior data quality to the SP survey which was administered by mail. Although the focus of their study is not the mode effects and the survey mode in comparison (mail) is different from our study (internet), their findings, compared with ours, indicate a possible effect of location setting. In our case, despite the fact that interviewer tried to intercept respondent at an opportune time to minimize disruption, respondents perhaps felt interrupted by the interview or too rushed to understand the idea of choice experiment and then shortcut the response process.[14]

Since our survey instruments were not designed for assessing the effects of different types of in-person interview or locations, it is therefore not possible for this paper to make a clear conclusion whether the larger variance of in-person interview SP data we found is caused by mode effects or other factors like the location setting. Nevertheless, given that most of the results in preference homogeneity tests of both in-person interview and internet SP datasets indicate that they more likely have the same underlying preference structure as the RP data, an only possible conclusion that can be drawn from our findings is that the pressure (from either interviewer or surroundings) felt by respondent during answering the questions more likely induces satisficing rather than social desirability bias.[15]

Conclusion

A number of findings are generated by this study. First, contrary to what is often assumed, the in-person interview does not always generate superior data quality. This study demonstrates that the underlying preference structure of the SP data collected from self-administered internet survey is not statistically different from that observed from the RP data collected from in-person interview on-site survey, whereas the underlying preference structure observed from the SP data – which was a part of the RP survey – is not always the same as that observed from the RP data. Second, variety of tests and measures for assessing the predictive ability

---

[14] The in-person interview SP data used in this study is also confounded with different timing of interview (i.e. as the respondents arrive, depart, or at any convenient time), which probably impacts the processing capability of the respondents. This is one of the promising areas for future research.

[15] We argue that a comparison of underlying preference structure of SP data with that of RP data can better represent the presence of social desirability bias than a comparison of derived welfare measures like marginal WTPs which is subsequently calculated from the respective preference structures.

reported in this study suggest that the in-person interview SP data is not superior to the internet SP data. Third, since the lack of the site-specific information of some attributes used in two SP models terribly affects their predictive ability, a comparison of predictive performances among two SP models and the RP model seems to be unwise. Although we cannot confidently say, as what is often assumed, that the RP model is superior to two SP models in modeling recreational demand, the RP model appears to predict the choices of the holdout samples quite well considering its most parsimonious model structure. The most reliable predictive performances of two joint SP-RP models nevertheless demonstrate the most promising alternative models in future study.

Although it is not possible for this paper to make a clear conclusion why the data quality of SP study administered in-person is not superior to that of self-administered internet survey, the findings of this study nevertheless suggest that a conventional belief of superiority of SP study administered in-person is not always true. Hence, with the caveat of confounding sample composition effects,[16] the study indicates a great potential of internet survey as an alternative survey mode for the hypothetical study of recreational demand.

The findings in this study raise some interesting issues for future research. In order to examine the *true* effects of different survey modes (in our case, in-person interview and internet surveys), some factors must be controlled to get better understanding of the differences in how respondents respond to the in-person interview SP surveys at different location. In our case, for instance, it would be preferable to isolate the influences of the location and timing settings by conducting the on-site interviews as the respondents arrive at two different locations, perhaps, one collected at a central facility (as Haener et al., 2001) and one collected at convenient spots, using the same survey instrument. Moreover, only the samples from the same region in both in-person interview and internet data sets are used in analysis in order to prevent the confounding effect of different the sample frames.

References

Adamowicz, W., Louvier, J, & Williams, M. (1994). Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. *Journal of Environmental Economics and Management, 26*, 271-292.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., & Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation. *Federal Register, 58*(10), 4601-4614.

Ben-Akiva, M. & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: MIT Press.

[16] On the other hand, Swait & Adamowicz (2001a) suggest that we should expect higher variance in preference for the complex choice tasks even if all respondents are identical. This occurs because some or all respondents are probably not providing sufficient effort in making the utility-maximizing choices, leading different respondents to make different choices.

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., & Rao, V. (1994). Combining Revealed and Stated Preferences Data. *Marketing Letters, 5*(4), 335-350.

Boyle, K. J., Morrison, M., MacDonald, D. H., Duncan, R., & Rose, J. (2016). Investigating Internet and Mail Implementation of Stated-Preference Surveys While Controlling for Differences in Sample Frames. *Environmental and Resource Economics, 64*(3), 401-419.

Bunch., D. S., Louviere, J. J., & Anderson, D. A. (1996). *A Comparison of Experimental Design Strategies for Choice-Based Conjoint Analysis with Generic-Attribute Multinomial Logit Models.* Working Paper, University of California, Davis, Graduate School of Management.

Carson, R. T. (2000). Contingent Valuation: A User's Guide. *Environmental Science & Technology, 34*(8), 1413-1418.

Collins, A., Rose, J.M. & Hess, S. (2012). Interactive Stated Choice Surveys: a Study of Air Travel Behaviour, . *Transportation, 39*(1), 55-79.

Department of National Parks, Wildlife and Plant Conservation (DNP). (2012). *Four Years Operation Plan (2012-2015).* Retrieved July 22, 2014, from http://www.dnp.go.th/info_office/dnpplan4y54.asp (in Thai)

Department of National Parks, Wildlife, and Plant Conservation (DNP). (2006). *Final Report: The Project to Study Carrying Capacity of Khao Yai National Park.* Retrieved November 21, 2013, from http://www.dnp.go.th/nprd/detailproject/cc_khaoyai.php (in Thai)

Haab, T. C., & McConnell, K. E. (2002). *Valuing Environmental and Natural Resources: the Econometrics of Non-market Valuation.* Massachusetts: Edward Elgar Publishing, Inc.

Haener, M. K., Boxall, P. C., & Adamowicz, W. L. (2001). Modeling Recreation Site Choice: Do Hypothetical Choices Reflect Actual Behavior? *American Journal of Agriculutral Economics, 83*(3), 629-642.

Hanemann, M. (1984). *Multiple Site Demand Models. Part II: Review of Existing Models and Development of New Models.* CUDARE Working Papers, UC Berkeley, Department of Agricultural and Resource Economics.

Heckathorn, D. D. (1977). Respondent-driven sampling: a new approach to the study of hidden populations. Social problems, 44(2), 174-199.

Heckathorn, D. D. (2011). Snowball Versus Respondent-driven Sampling. Sociological Methodology, 41(1), 355-366.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied Choice Analysis: A Primer.* New York: Cambridge University Press.

Hensher, D., & Bradley, M. (1993). Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models. Marketing Letters, 4(2), 139-151.

Horowitz, J. L., & Louviere, J. J. (1993). Testing Predicted Choices against Observations in Probabilistic Discrete-Choice Models. *Marketing Science, 12*(3), 270-279.

Kamolthip, S. (2016). *Enhancing the Effectiveness of the Policy Design of Thailand's National Parks Using Random Utility Model: Case Study of Khao Yai National Park (Unpublished Doctoral Dissertation).* National Institute of Development Administration, Bangkok.

Kuhfeld, W. (2010). *Marketing Research Methods in SAS: Experimental Design, Choice, Conjoint, and Graphical Techniques*. Retrieved October 7, 2014, from https://support.sas.com/techsup/technote/ mr2010.pdf

Lindhjem, H., & Navrud, S. (2011a). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics, 70*(9), 1628-1637.

Lindhjem, H., & Navrud, S. (2011b). Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes. *International Review of Environmental and Resource Economics, 5*, 309-351.

Louviere, J. J. (1988). *Analyzing Decision Making: Metric Conjoint Analysis (Sage Uviversity Paper Series on Quantitative Applications in the Social Sciences, series no. 07-067).* Newbury Park, CA: SAGE Publications, Inc.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated Choice Methods: Analysis and Application.* New York: Cambridge University Press.

Manski, C. F., & Lerman, S. R. (1977). The Estimation of Choice Probabilities from Choice Based Samples. *Econometrica, 45*(8), 1977-1988.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

McFadden, D. (1996). *On the Analysis of "Intercept and Follow" Surveys.* Working Paper, UC Berkeley.

Miller, K. M., Hofstetter, R., Krohmer, H., & Zhang, Z. J. (2011). How Should Consumers' Willingness to Pay Be Measured? An Empirical Comparison of State-of-the-Art Approaches. *Journal of Marketing Research, 48*(1), 172-184.

Swait, J. & Adamowicz, W. (2001a). Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. *Organizational Behavior and Human Decision Processes, 86*(2), 141-167.

Swait, J. & Adamowicz, W. (2001b). The Influence of Task Complexity onConsumer Choice: A Latent Class Model ofDecision Strategy Switching. *Journal of Consumer Research, 28*(1), 135-148.

Swait, J. & Louvier, J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research, 30*(3), 305-314.

Tanakanjana, N., Arunpraparut, W., Pongpattananurak, N., Nuampukdee, R., & Chumsangsri, T. (2006). *The Project "Decision Support System for Sustainable Management Planning of Nature-based Recreation Areas, Phase I".* Reserach Report, Thailand Research Fund (TRF), Bangkok. (in Thai)

Tanakanjana, N., Phumsathan, S., & Nunsong, N. (2012). *Sustainable Uses of Natural Resources in Protected Areas for Recreation and Tourism.* Research Report, National Research Council of Thailand (NRCT), Bangkok. (in Thai)

Whittington, D. (1998). *Administering Contingent Valuation Surveys in Developing Countries*. World Development, *26*(1), 21-30.