RURAL ECONOMY

Modeling Recreation Site Choice: Do Hypothetical Choices Reflect Actual Behavior?

M. Haener, P.C. Boxall and W.L. Adamowicz

Staff Paper 00-01

Staff Paper

Department of Rural Economy
Faculty of Agriculture, Forestry
and Home Economics
University of Alberta
Edmonton, Canada

Modeling Recreation Site Choice: Do Hypothetical Choices Reflect Actual Behavior?

M. Haener, P.C. Boxall and W.L. Adamowicz

Staff Paper 00-01

Please Do Not Cite or Quote Without Permission

The authors are, respectively, Research Associate, Associate Professor and Professor, Department of Rural Economy, University of Alberta, Edmonton Alberta.

The purpose of the Rural Economy 'Staff Papers' series is to provide a forum to accelerate the presentation of issues, concepts, ideas and research results within the academic and professional community. Staff Papers are published without peer review.

This research was funded by Mistik Management, the Canada-Alberta Partnership Agreement in Forestry, the Canada-Saskatchewan Partnership Agreement in Forestry, and the Canadian Forest Service.

This study examines the ability of revealed preference (RP), site-specific stated preference (SP), transferred SP, and various joint RP-SP models to predict aggregate and individual recreation site choice behavior in a holdout sample. For two statistical comparisons, the site-specific RP model provided the most accurate predictions of individual choices. However, the transferred SP model, applied directly or estimated jointly with the RP data, performed best in three aggregate and one individual prediction tests and second best in the other individual prediction comparisons. In every test examined the transferred SP models outperformed the site-specific SP models. This result is traced to the method used to collect the hypothetical choice data (mail out vs. in-person settings) and illustrates the importance of data quality in accuracy of behavioral prediction. These findings suggest that data from well designed and conducted SP surveys from one site can be combined with site-specific RP data from another site to generate improved models of recreation site choice.

Key words: choice experiments, nonmarket valuation, prediction of behavior, recreation site choice.

Do Hypothetical Choices Reflect Actual Behavior? A Comparison of the Predictive Abilities RP and SP Models

An objective of many economic studies is the ability to predict behavior. Model predictions can inform policy makers of the probable results of changes in resource management or demography. However, developing models with predictive power is a formidable challenge within the social sciences.

Environmental economists have used probabilistic models such as the random utility model to build models of recreation site choice. Such models can be based on actual or revealed behavior (revealed preference- RP), or on responses to hypothetical scenarios (stated preference- SP). There is not complete agreement as to whether RP or SP data are best at producing models which predict behavior (Louviere 1994; Blamey et al. 1999b). It is often assumed that models based on actual behavior have superior predictive ability compared to models based on hypothetical behavior. However, it is also recognized that as a result of their controlled design, SP models are desirable because they capture responses to changes in attributes (Swait, Louviere and Williams.). Therefore, as noted by Louviere (1994:19), it is an open empirical question whether SP models estimated from choice experiment data predict behavior more accurately than RP models estimated from data in real markets.

In recent years, many advances were made in modeling behavioral choice, particularly through the use of random utility models. These advances include the development of joint or sequential estimation of RP and SP data which in theory allows the strengths of each data type to be realized in developing superior choice models. Several studies have shown that the *within sample* predictive ability of joint models exceeds that of models estimated with only RP data (Adamowicz et al.; Swait, Louviere and Williams).

Recent research has also focussed on the transferability of preferences across regions and/or time. In most studies model or equation transfers were found to be more feasible and defensible than the transfer of point estimates of benefits (Bergstrom and De Civita). Many different tests and measures have been proposed to assess the transferability of discrete choice models, particularly in the transportation literature (see Koppelman and Wilmot, Ortuzar and Williamson). Some of these tests include formal tests of parameter equality (within a factor of proportionality) using the techniques of Swait and Louviere. Predictive ability has also been used as a means of assessing the transferability of discrete choice models (Atherton and Ben-

Akiva). In the environmental economic literature, however, only Parsons and Kealy use predictive ability as a test of benefits transfer. In their empirical example they suggest that if a model from one region accurately predicts behavior in another region, it could be used to conduct benefits transfers between the regions.

We possess discrete choice RP data and SP data for the same activity from two different regions (a total of 4 data sources). This provided us with the opportunity to compare the prediction success of RP, SP and joint RP-SP models, and to use this information to examine the transferability of SP models across these two regions. In this examination we focus on the ability of these models to predict holdout sample choices since we feel that this aspect of model performance is often neglected. We agree with Horowitz and Louviere that models that cannot satisfactorily predict choices in holdout samples are less useful than those that can. Our motives are driven by the need to develop predictive models to be used as decision support tools for resource management, and to expand current criteria used to examine the potential for benefits transfer.

In this paper we examine the "performance" of several econometric models and data generating mechanisms. Performance is measured in terms of the success in aggregate and individual prediction of a holdout sample. In our analysis, we employ relatively simple econometric models. Prediction success could potentially be improved by considering heterogeneity (with random parameter or latent class models), correlation structure (nested models or multinomial probit models), temporal dimensions (habits, etc.) or a variety of other factors. However, we wish to examine prediction success, and the performance of various data generating mechanisms, using common and simple approaches. We also wish to employ techniques that practitioners are able to implement using currently available software. The results suggest that even with these simple models, some data types and modeling strategies produce remarkably accurate aggregate predictions. Furthermore, distinct differences in performance can be detected even in these relatively simple models.

Background and Motivation

A number of studies examined recreational moose hunting in Canada (Morton et al. 1995; Boxall et al. 1996; Adamowicz et al. 1997; Boxall and McNab 1999). In two of these studies choice experiments involving hypothetical sites and information on hunting trips to actual sites was collected from two different geographical areas (west-central Alberta and central Saskatchewan). Thus, two choice experiments (SP data) and two RP data sets exist. These datasets provided an opportunity for us to examine statistical differences in preferences over the attributes involved, and the success of some single and joint models in predicting behavioral site choice between these two geographical areas.

Our interest in this endeavor was motivated by the need of forest managers in Saskatchewan to understand the impact of forestry operations on moose hunting. In this case, the managers did not have current information on moose hunters in their forest management area in northwestern Saskatchewan. However, the availability of Alberta data and Saskatchewan data from an adjacent forest management area in central Alberta, prompted us to examine opportunities for benefits transfer. Using the data available we can assess whether hunting preferences differ significantly by region. This information will help us determine whether we need site-specific information for northwestern Saskatchewan or whether models from other boreal forest region would be applicable. To examine the transferability of preferences, we developed a holdout sample from the central Saskatchewan survey data. This sample allows the examination of the success of an RP model estimated on information from those not in the holdout sample, but who actually visited the area. Our expectation at the outset was that the performance of this RP model would be the benchmark to compare other site choice models with.

In the marketing and transportation literatures assessing the success with which SP models predict actual behavior is used as a test of transferability. For example, Koppleman and Wilmot examine prediction success and transferability in a transportation setting, and Horowitz and Louviere assess prediction success for college choice by students.

In the environmental economics literature, there have been a variety of tests of transferability (e.g. Bergstrom and de Civita) but most have used comparisons of welfare measures or model parameters as a measure of the success of transferability. While this approach does provide insight into the transferability of models, it does not provide an overall impression of the ability of these models to predict behavior. The coefficients estimated from models for two regions may be "similar", but the models may not predict actual choices very well.

In a recent paper, Blamey et al. (1999b) compare the results of two SP choice models with RP data for a green product (toilet paper). Blamey et al. (1999b:3) point out that divergences in RP and SP estimates may result from "measurement error associated with the inclusion of objective attribute definition and levels in SP questionnaires rather than those perceived by respondents." Blamey et al. (1999b) also note that a stronger test of predictive validity would involve the assessment of prediction success for choice sets different to those represented in the RP data set. We essentially conduct this type of assessment in this study. We use the information from SP, RP and joint SP-RP models, estimated from different data sources, to predict a holdout sample of actual choices. The approach provides insight into the ability of SP, RP and joint models to predict actual choice.

Data and Models

The Saskatchewan Revealed Preference Data

In 1994, the Canadian Forest Service conducted a moose hunting survey involving a specific area: the Weyerhaeuser Forest Management Lease Area (FMLA) in central Saskatchewan. Residents of larger centers in the vicinity of the FMLA were included in the sample and residents of smaller communities in the area were over-sampled.

In the survey respondents were asked to record the frequency and location (wildlife management zone (WMZ)) of hunting trips they took in 1993. They were also asked to indicate the level of access, encounters, forestry activity and moose populations that best described the 11 WMZs in the study region. Similar information was also elicited from wildlife biologists and conservation officers from the region. These responses in combination with information related to hunting activity and wildlife populations were used to create an attribute matrix that described the WMZs in the region.

The survey was mailed to 1274 individuals whose names and addresses were obtained from registered licensed hunters held by the Government of Saskatchewan. The Saskatchewan sample collected information from about 660 respondents. A total of 706 trips were taken to the WMZs included in the study region by 315 individuals. The holdout sample consisted of about half these people (157 individuals) randomly drawn from the respondents. The choices of the remaining individuals were used to estimate the RP models reported below.

The Alberta and Saskatchewan Stated Preference Data

The 1993 Alberta moose hunting choice experiment is described by Boxall et al. and Adamowicz et al. In this experiment six measurable attributes associated with moose hunting experiences (Table 1) of either 2 or 4 levels were determined. Experimental design methods (see Louviere 1988) were used to produce 32 choice sets that were blocked into two sets of 16. A hunter sampled in the study was presented with one of these 16 pairs of alternative descriptions[1] of moose hunting sites. This choice experiment was administered to samples of hunters selected from Alberta Fish and Wildlife Services license records. The hunters were sampled from 5 locations, four located in within the study area in west central Alberta and the fifth a large metropolitan centre (Edmonton) located about 100 km outside the area. Telephone recruitment and reminders generated a sample of 271 hunters. The choice experiment instruments were administered in person to groups of hunters ranging from 20 to 55 individuals at 8 meetings held in various locations throughout the study area. The final data set contained information from 271 respondents who provided answers for 4080 choice scenarios.

The 1994 Saskatchewan moose hunting mail survey also included a choice experiment that contained a similar presentation to the one used in the Alberta study. However, in the Saskatchewan study the Road Quality attribute was deleted and a Wildlife Species Diversity attribute was added. The attributes were also described in 3 levels rather than 2 or 4. The resulting experimental design generated 2 versions of 14 choice scenarios (see Boxall and McNab). The Saskatchewan sample involved information from about 660 hunters who provided answers for about 7832 choice scenarios. In order to compare this information with the Alberta data, a random sample of 4080 choices was drawn from this data to ensure comparability with the Alberta sample.

Although the two choice experiment designs are not exactly the same, there is a large degree of similarity in the attributes that were included in the designs. To improve comparability, three of the Alberta attributes were coded to match the levels in the Saskatchewan

[1] The option of not choosing either alternative (not going moose hunting) was also presented with each pair.

data, and one Saskatchewan attribute (Forestry Activity) was coded to match the Alberta attribute (see Table 1). The coding permitted the attributes common between the two datasets to have the same number of levels. These levels also correspond more closely to the information about the WMZs gathered from the Saskatchewan wildlife biologists and conservation officers.

Socioeconomic data were also collected as part of both the Alberta and Saskatchewan surveys. Residence location was elicited and used to categorize the respondent as rural or urban. In the Alberta sample only participants from Edmonton were included in the urban category since the other respondents were from communities of less than 25,000 people. In the Saskatchewan sample, urban respondents included those from several communities with populations exceeding 25,000. Other demographic information collected includes income, age, education level and years of hunting experience[2].

In the models presented below, we examine the frequency of alternatives chosen from a fixed choice set. The choice set for the SP data is a pair of designed, hypothetical hunting sites and an option to not go hunting. The choice set for the Saskatchewan RP data is the set of 11 WMZs in central Saskatchewan. The models developed from these data sets are then used to predict choices over the 11 WMZs made by the holdout sample.

Theory and Econometric Estimation of the Choice Models

We utilize the discrete choice model to analyze the choice between alternative recreation sites. The discrete choice model is based on random utility theory that postulates that an individual (in this case, a hunter) will select the option that provides them with the greatest utility. Therefore, the probability of selecting an alternative increases as the utility associated with it increases. The utility that an individual derives from an alternative is considered to be associated with the attributes of the alternative. The utility function is composed of a deterministic component (*V*) and an unobservable or stochastic component (ε):

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

V is the indirect utility function in which the attributes are arguments. Therefore, *V* can be characterized as:

(2) $$V_i = \beta_k X_i$$

where X is a vector of k attributes associated with alternative i and β is a coefficient vector. If the distribution of the stochastic component or error terms, is characterized as IID Gumbel, McFadden shows that the conditional choice probability of selecting alternative i is:

(3) $$prob(i) = \exp(\mu\beta_k X_i) \Big/ \sum_{j \in C} \exp(\mu\beta_k X_j)$$

Where μ is a scale parameter and C is the choice set. When a single set of data is used to estimate a model, μ is confounded with the parameter vector and cannot be identified. When estimating the RP and SP models reported in this study, we assume $\mu=1$ and the parameters are estimated using maximum likelihood methods.

If two complementary samples are jointly estimated, however, then the ratio of scale parameters can be determined. This would mean that the parameter vectors between the groups or samples differ by a scale or factor of proportionality. However, since the scale parameter is inversely proportional to error variance, it also means that the two samples display different levels of error in their choices. Equation 4 below (from Louviere, Hensher and Swait) illustrates the relationship between the relative scale and variance for two samples of data where variance is represented by σ, scale by μ and the two different samples are designated A and B.

(4) $$\frac{\sigma_A^2}{\sigma_B^2} = \frac{\pi^2/6\mu_A^2}{\pi^2/6\mu_B^2} = \frac{\mu_B^2}{\mu_A^2} = \left(\frac{\mu_B}{\mu_A}\right)^2$$

The ratio of scale parameters becomes relevant where joint models are estimated using data consisting of both RP and SP observations, which we do in this study. Following Adamowicz *et al.* (1997), the likelihood function we used for these joint conditional logit models was:

(5) $$L(\beta, Z^{RP}, Z^{SP}, \tau) = \sum_{n=1}^{N^{RP}} \sum_{i \in C_n} f_{in}^{RP} \ln \Pr\left\{i \mid \beta, Z^{RP}\right\} + \sum_{n=1}^{N^{SP}} \sum_{i \in C_n} f_{in}^{SP} \ln \Pr\left\{i \mid \beta, Z^{SP}, \tau\right\}$$

where n indexes individuals from the RP and SP samples; i indexes alternatives; f_{in}^{RP}, f_{in}^{SP} are the frequencies of choice in the RP and SP observations, respectively; $\Pr\left\{i \mid \beta, Z^{RP}\right\}$

[2] More information related to the Saskatchewan survey is provided in McFarlane and MacNab (1999). The Alberta survey is described in detail by McLeod et al.

and $\Pr\left\{i \mid \beta, Z^{SP}, \tau\right\}$ are the probabilities of an individual *n* choosing alternative *i* in the RP and SP samples, respectively; β is the parameter vector common between the RP and SP data which is restricted to be equal in estimation; Z^{SP} and Z^{RP} are parameter vectors associated with variables unique to the RP and SP data; and τ represents μ_{SP}/μ_{RP}, or the ratio of the scale of the SP data to that of the RP data. As in Adamowicz et al., the choice variable for the RP data, f_{in}^{RP}, is specified as proportions so as to eliminate the possible over-weighting of RP observations in the joint model.

Joint estimation was achieved by vertically concatenating the data matrices of the individual data sets and estimating a single set of parameters. For variables common between the data sets, the coefficients are restricted to be equal within a factor of proportionality or scale. The attribute matrix of one data set is multiplied by this relative scale parameter. The estimation of this relative scale parameter allows variance differences between the data sets to be accounted for in the estimation procedure (see Swait and Louviere).

Parameter Estimates

In estimating model parameters, the distance variable was expressed as travel costs[3] and the other attributes and levels of the choice alternatives were effects coded identically across all the RP and SP models. Note that effects coded attributes result in one fewer parameter than the number of levels; thus for a 3-level attribute, the coefficient on the 3rd level is the negative sum of the coefficients on the other 2 levels (see Louviere 1988; Boxall and MacNab). The coefficient vectors, log likelihood values, and ρ^2 statistics for 10 models are reported in Table 2. The data sources used for each model are described by the labels listed below:

[3] In this calculation we value out-of-pocket expenses at $0.28/km and the opportunity cost of time was estimated as one third of an individual's hourly wage (income/2040 hrs) and an assumed speed of 80 km/hr.

Acronym	Data Used in Estimation
SK-RP1	Saskatchewan RP data, Specification 1
SK-RP2	Saskatchewan RP data, Specification 2
SK-SP1	Saskatchewan SP data, Specification 1
SK-SP2	Saskatchewan SP data, Specification 1
AB-SP1	Alberta SP data, Specification 1
AB-SP2	Alberta SP data, Specification 2
J-ABSP1	Joint Alberta SP, Saskatchewan RP, Specification 1
J-ABSP2	Joint Alberta SP, Saskatchewan RP, Specification 2
J-SKSP1	Joint Saskatchewan SP, Saskatchewan RP, Specification 1
J-SKSP2	Joint Saskatchewan SP, Saskatchewan RP, Specification 2

The Saskatchewan RP data (non-holdout sub-sample) contained information from the sub-sample of 370 trips to the study region from 157 hunters. The levels of the attributes at the 11 WMZs in the region were developed from the expert judgements of regional biologists and foresters working in the area. The hunters' choices were represented as proportions in the econometric estimations. The two RP models estimated differed in that the second model (SK-RP2) included an urban travel cost interaction term (Table 2). This term was included to facilitate comparison with the SP and joint models described below. The signs of the parameters are consistent with theory and previous research on moose hunting. For example, travel costs are negative, WMZs with fewer encounters are preferred, and zones with higher moose encounters are desired. Note that coefficients could not be estimated for all attributes due to lack of variation in their levels at the 11 sites (based on expert judgements), and that where only one is reported, the other level is simply the negative of the reported coefficient.

Two Saskatchewan SP models were estimated: one without and one with an urban travel cost interaction term (Table 2 columns 4-5). The parameter vectors show that the patterns of preferences across the attributes are similar to those observed in the RP models, but the weights on the attributes are quite different. The wildlife viewing attribute is statistically significant in these models, but not in the RP models.

The Alberta SP data was used to estimate two models (see Table 2). Due to the slightly different design of the AB choice experiment, the parameter vector differs from the Saskatchewan models. A full set of urban interaction terms was included in the second model (AB-SP2). These interaction terms are all significant at the 90% level of confidence, but due to space limitations are not reported. The patterns of preferences implied by the parameters is similar to both the RP and SP models estimated from Saskatchewan data.

The final set of models consists of four joint models that utilize both RP and SP data in estimation (see Table 2). Since the data now contain RP information, some parameters not identified in the Alberta SP models could be estimated (e.g. the wildlife viewing attributes). The parameters of these joint models are consistent with their SP counterparts, but the magnitudes of those parameters common between the RP and SP data change. In other words, the inclusion of the RP information results in different weights on the attributes.

We did not include alternative specific constants (ASCs) for WMZs in the parameter vectors for any models that used RP information. ASCs are typically included to capture the utility of an alternative that is not captured by the attributes in the model. A complete set of ASCs would produce prefect within sample prediction success. ASCs are thought to generally improve model performance, but they cannot be used in predicting the effect of changes due to attribute changes. Ideally, one would want to use attributes to thoroughly explain choice (Adamowicz et al.1997:73). Furthermore, unless one employs a “branded” SP experiment (i.e. Blamey et al. 1999a), ASCs relevant to the actual sites cannot be determined from SP data. In order to construct a “fair” comparison between SP and RP data we did not include ASCs in the RP or joint models.

Tests of Model Performance: Prediction Tests

A variety of tests and measures have been used to compare the predictive ability of choice models. Many of these tests operate at the aggregate level comparing observed and predicted market shares or in this case observed and predicted trip distribution. However, there are also tests that utilize prediction success at the individual level. We utilize both in comparing the abilities of the 10 models to predict the choices of the holdout sample.

Koppelman and Wilmot review some measures of predictive ability based on aggregate predictions. The first is the sum of absolute errors (SAE) which gives equal weight to all errors. The calculation of the SAE is simply:

(6) $$SAE = \sum_{i=1}^{J} (\hat{N}_i - N_i)$$

where $\hat{N}_i$ is the total number of trips predicted to destination *i*, N_i is the number of trips observed to destination *i* and *J* is the number of observations.

Another common measure is the aggregate prediction statistic (APS) which is generated from a one sample $\mathbf{P}^2$ test of the hypothesis that the observed frequencies of choice in each group are collectively generated by the prediction model (Koppelman and Wilmot). In our case, we did not aggregate destinations; therefore each of the 11 sites represents a group as defined by Koppelman and Wilmot. Following Siegel, the one sample $\mathbf{P}^2$ statistic is calculated as:

(7) $$\chi^2 = \sum_{i=1}^{J} (\hat{N}_i - N_i)^2 \Big/ N_i$$

with degrees of freedom *J*-1. Koppelman and Wilmot (1982:20) note that this test assumes that the trip distribution is predicted without sampling error and as a result “is more likely to reject the hypothesis that all frequencies come from the candidate model than would a statistic that takes account of sampling variation”. We report this measure because many other studies use this statistic, but we demonstrate that it can provide some erroneous conclusions regarding prediction success.

The $\mathbf{P}^2$ statistic provides a measure of the weighted sum of squared errors. According to Siegel (1956:43), the larger the χ^2 is, the more likely it is that the observed frequencies did not come from the population upon which the null hypothesis is based. Due to the squaring of errors, the APS gives greater weight as the difference between the predicted and the actual frequency increases.

Horowitz and Louviere list other aggregate level tests of predictive ability. These include the degree of correlation between predicted and observed market shares, and regression tests for a slope of one and intercept of zero in a regression of observed aggregate shares on predicted shares. Since these measures would result in the same ordinal ranking of models, we only computed the degree of correlation between the predicted and observed aggregate distributions.

The tests described above are based on assessment of aggregate prediction success. However, tests of predictive ability at the individual level may be more useful. These tests operate by comparing the observed and predicted choices for each individual in the sample. Horowitz and Louviere developed a test that involves regressing observed shares (probabilities) on predicted shares (probabilities). However, they adjust this test to account for the effects of random sampling errors by including the variance-covariance matrix associated with the parameter estimates. Since we were only looking to compare models, we computed an overall correlation coefficient using the observed and predicted probabilities for each individual. We also computed individual-specific correlation coefficients (i.e. one for each of the individuals in the sample) and examine the distributions generated by this procedure.

Our final measure of individual level prediction success is McFadden's prediction success index (σ). This index is calculated from a prediction success table where the proportions of successful predictions for each alternative within the choice set are examined (Maddala). The index is then calculated as:

(8) $$\sigma = \sum_{i=1}^{m}\left[\frac{N_{ii}}{N..} - \left(\frac{N_{.i}}{N..}\right)^2\right]$$

where $N..$ is the total number of choice occasions, N_{ii} is the number of correct predictions for alternative i, $N_{.i}$ is the total number of choice occasions where the choice is predicted to be i.

Maddala notes that $\sigma > 0$ and that the maximum value of σ ($\sigma_{\max}$) is $1 - \sum_{i=1}^{m}\left(\frac{N.i}{N..}\right)^2$. We report the normalized index (σ_n) which involves dividing σ by $\sigma_{\max}$.

Comparison of Predicted Trip Distributions

The predictive abilities of the ten models were compared using the holdout sample of 157 of the SK-RP respondents. According to their reports, these individuals took a total of 336 hunting trips to the 11 wildlife management zones in the study region in 1993. The distribution of these trips is shown in Table 3. WMZs 59 and 63 were by far the most popular choices.

Each model in Table 2 was used to predict the trips taken by the individuals in the holdout sample. The parameter vectors of these models were combined with the site attributes generated through expert judgements and travel costs to estimate the probability of each individual visiting each WMZ. These probabilities were then used to determine how the total number of trips (336) were distributed across the sites.

The resulting aggregate predicted trip distributions are shown in Table 3 and for the RP and SP models the actual and predicted trip distributions are plotted in Figure 1. The RP models predict the holdout sample trips reasonably well. However, the Saskatchewan SP models predict poorly, particularly for WMZ 63. The Alberta models appear to predict actual trips more accurately than the SK SP models.

To further compare these distributions, the aggregate trip distributions were used to calculate the aggregate $\mathbf{P}^2$, the SAE, and the correlation coefficient between the observed shares and the predicted shares. These statistics are reported in the last three columns of Table 3. The J-ABSP and the RP models have the lowest $\mathbf{P}^2$ values, although their magnitudes suggest that the observed and predicted distributions are significantly different. The SAE values, on the other hand, suggest that the two Alberta SP models and one joint model using Alberta SP information had the lowest errors in prediction. Finally, the highest degree of correlation (largest correlation coefficient) between the aggregate observed and predicted trip distributions was associated with the Alberta SP models.

The results of the individual prediction tests are shown in Table 4. The overall correlation coefficient for individual (as opposed to aggregate) choices further highlights the higher accuracy of prediction of the RP models and the Alberta SP models. Those joint models using Saskatchewan SP information do not predict as well as those using the Alberta SP information. The other individual level statistics further suggest that the prediction accuracy of the RP models, the Alberta SP models, and the joint models using the Alberta SP information is higher than that of the Saskatchewan models. However, the normalized McFadden indices point to the Alberta SP models as the best predictors.

Individual specific correlation coefficients were calculated for each of the 157 individuals in the holdout sample using the predictions from each model. The means of the resulting distributions of correlation coefficients do not differ appreciably among the 10 models (Table 4). To further examine this information, histograms showing the distributions of these coefficients for one model in each group are shown in Figure 2. What is striking about these distributions is the biomodal nature of the distribution of Saskatchewan SP correlations; most of the coefficients are between 0.9 and 1.0 or close to 0. This pattern is not evident for the other three models.

The 10 models were ranked according to their prediction performance for each of the tests. The models that achieved the top five ranks are presented in Table 5. Note that while each test provides a different ranking of the models, none of the models using the Saskatchewan SP data make the top five. The aggregate level tests suggest that the J-ABSP2 and the AB-SP1 model are superior models. However, the χ^2 tests rate the Saskatchewan RP models high. The individual level tests provide somewhat different rankings than the aggregate tests. The overall correlation coefficient between predicted and actual choices suggests that the SK-RP2 model performs best, but both of the AB-SP models achieve good results as well. The mean of the individual correlations and the σ_n's also point to the performance of the RP models. However, the models using the Alberta SP information alone or jointly with the RP data predict the holdout sample's actual trips almost as well.

Discussion

These findings are unexpected in a number of ways. First, we expected the Saskatchewan RP models to outperform either the single or joint SP models in prediction success. This expectation was based on the belief in the literature that models based on actual behavior would predict actual behavior well. While some of our comparisons between the RP and SP predictions identify greater accuracy with the RP data (e.g. overall correlation coefficient for individual choices, Table 4), the overall picture of the success of the RP models is not clear.

Second, we reveal the rather surprising result that SP information from Albertan hunters who hunted in areas over XXX km away, predicts the behavior of the holdout sample far better than the SP information generated from individuals who actually hunted in the same area as the holdout sample. We expected the Saskatchewan SP models to out-perform the Alberta SP models. Of the single data set models, the SK-RP and AB-SP models perform best. Considering this, it is not surprising that a joint SK-RP and AB-SP model (J-ABSP2) generates the least predictive error of the joint models.

Explaining the prediction success of this Alberta data is challenging. One possible reason may be the difference of the degree of error variance between the two sources of SP data. Information on this can be gained by examining the relative scale parameter in the joint models.

In these models, the scale of the SK-RP data was normalized to 1. Because of the relationship between the relative scale parameter and the ratio of error variances (equation 4) above we can compare the relative error variance in the SK-SP and AB-SP data. From J-ABSK2 model the relative scale ratio ($\mu_{SK\text{-}SP}/\mu_{RP}$) is about 0.6 (Table 2) and using equation (4), the variance of the SK-RP data is about 36% that of the SK-SP data. Similarly from the joint estimation of the AB-SP/SK-RP models (J-ABSP2) we attain a scale ratio of about 0.68, indicating that the variance of the SK-RP data is about 46% of the variance of the AB-SP data. This identifies that the error variance of the SK-SP data is about 1.3 times that of the AB-SP data.

The difference in error variance between the SP data sets could be due to differences in the administration of the survey instruments. The Saskatchewan SP survey was administered by mail, whereas the Alberta SP survey was administered in-person during group meetings. The in-person setting of the AB survey may have generated more thought-out and reliable responses. In addition, the research team was present to explain the survey and provide answers to questions. This method of data collection may have led to superior data quality and this might have led to the development of superior choice models. We believe that this difference in data collection methods between the two SP studies is a major explanator of the differences in the success of predicting the behavior of the holdout sample.

Our results suggest that under some conditions SP models can perform as well or better than RP models. This conclusion is contrary to those of Blamey et al. (1999), who find that RP data generates models with better predictive ability than SP models. However, their test is within-sample and is restricted to aggregate level prediction. In our case the comparison involves a holdout sample, and we use a much wider array of aggregate and individual level prediction tests. The quality of the SP data may explain prediction performance in cases where SP models predict poorly compared to RP models for the same area. Our results suggest that carefully designed and implemented SP studies can generate models that predict actual behavior as well or better than RP models. However, even greater benefits may be realized from a well-conducted SP survey because, as shown here, it may be transferable to other regions and used in combination with site-specific RP data to generate useful joint models.

Conclusions

This paper has generated a number of findings. First we demonstrate that, contrary to what is often assumed, SP surveys can be used to estimate models that have predictive ability on par with RP models. These SP models appear to predict the choices of a holdout sample of data quite well, a result that is comforting to users of such approaches. Unfortunately, some of the models, in this case even ones based on data collected in the region, did not predict choices well. Second, if the data quality is superior, it may be preferable to use a transferred SP model instead of a site-specific SP model to predict aggregate behavior (trip distributions) in the site of interest. Third, we find that combining site-specific RP data with transferred SP data generates a joint model with the best superior predictive ability.

From our findings we can draw a number of conclusions. It appears that surveys conducted in person or via group session where the interviewer is present attain superior data quality relative to mail surveys. Although in person or group administration may be more expensive the investment may be warranted given that data quality will be improved. Further, fewer site-specific surveys would be warranted, since the data could be combined with site-specific RP data in other regions to generate joint models. The development of a few data collection efforts of high quality may provide data useful for benefit transfers. The NOAA panel (Arrow et al.) made a similar proposal when they recommended the development of some benchmark surveys of passive use values.

These findings raise several issues requiring further research. Understanding the benefits of in-person surveys compared to mail-out surveys and the differences in how individuals respond to these methods would contribute to the development of better SP studies. In pursuit of more realistic predictive models, we also need to demonstrate how well RP, SP and joint models predict behavioral responses of actual (real life) attribute changes. The results reported in this paper, in conjunction with other current research (e.g. Swait, Louviere and Williams), suggest that SP models may be best at predicting responses of individuals facing resource tradeoffs. This suggestion needs to be tested in other studies involving different activities and environments.

References

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, M. Williams. "Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation." *J. Environ. Econ. and Manage.* 32(1997):65-84.

Arrow, K., R. Solow, P. Portnoy, E. Leamer, R. Radner and H. Schuman. Report of the NOAA Panel on Contingent Valuation. *Fed. Reg.* 4601-4614, (1993).

Atherton, T. J. and M. E. Ben-Akiva. "Transferability and Updating of Disaggregate Travel Demand Models". *Transp. Res. Rec.* 610(1976):12-18.

Bergstrom, J. C. and P. De Civita. "Status of Benefits Transfer in the United States and Canada: A Review." *Can . J. Agr. Econ.* 47(1999):79-97.

Blamey, R. K., J.W. Bennett, J.J. Louviere, M.D. Morrison and J. C. Rolfe. "The Use of Policy Labels in Environmental Choice Modeling Studies." Choice modeling research report No. 9, School of Economics and Management, University College. Canberra, Australia: University of New South Wales, 1999a.

Blamey, R. K., J.W. Bennett, J.J. Louviere, and M.D. Morrison. "Validation of a choice experiment involving green product choice." Choice modeling research report No. 10, School of Economics and Management, University College. Canberra, Australia: University of New South Wales, 1999b.

Boxall, P.C., W.L. Adamowicz, J. Swait, M. Williams, and J.J. Louviere. "A Comparison of Stated Preference Methods for Environmental Valuation." *Ecol. Econ.* 18(1996):243-253.

Boxall, P.C. and B. MacNab. "Exploring the Preferences of Wildlife Recreationists for Features of Boreal Forest Management: A Choice Experiment Approach." Unpublished manuscript under review, 1999.

Horowitz, J.L. and J.J. Louviere. "Testing Predicted Choices Against Observations in Probabilistic Discrete Choice Models." *Mktg. Sci.* 12(1993): 270-279.

Koppelman, F. S. and C. G. Wilmot. "Transferability Analysis of Disaggregate Choice Models." *Transp. Res. Rec.* 895(1982):18-24.

Louviere, J.J. *Analyzing Decision Making: Metric Conjoint Analysis*. Sage University Papers Series No. 67. Newbury Park CA: Sage Publications, 1988.

Louviere, J. "Combining Revealed and Stated Preferences Data: The Rescaling Revolution." Paper prepared for Association of Environmental and Resource Economists, Lake Tahoe, 2 June 1996.

Louviere, J.J. "Relating Stated Preference Measures and Models to Choices in Real Markets: Calibration of CV Responses." Paper prepared for the DOE/EPA Workshop on Using Contingent Valuation to Measure Non-Market Values. Herndon VA, 19-20 May 1994.

Louviere, J.J., D. Hensher, and J. Swait. Choice Modeling Book, 1999.

Maddala, G.S. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, 1983.

McFarlane, B.L. and B. MacNab. Human Dimensions of the Saskatchewan Forest Habitat Project. Information Report (forthcoming), Northern Forestry Centre, Canadian Forest Service, Edmonton, Alberta, Canada, 1999.

McLeod, K, P.C. Boxall, W.L. Adamowicz, W.L., M. Williams, and J.J. Louviere. Incorporation of non-timber goods and services in integrated resource management. Project Report Number 93-12, Dept. Rural Economy, University of Alberta, Edmonton, Alberta, 1993.

Morrison, M. Choice Modeling Non-Use Values, and Benefit Transfer. PhD Thesis. University of New South Wales, Canberra, Australia, 1998.

Morton, K.M., W.L. Adamowicz and P.C. Boxall. "Economic Effects of Environmental Quality Change on Recreational Hunting in Northwestern Saskatchewan: A Contingent Behaviour Analysis." *Can. J. For. Res.* 25(1995):912-920.

Ortuzar, J. and L. G. Williamson. *Modeling Transport*. New York: J. Wiley and Sons, 1990.

Parsons, G.R and M.J Kealy. "Benefits Transfer in a Random Utility Model of Recreation." *Water Resour. Res.* 30(August 1994):2477-2484.

Siegel, S. *Nonparametric Statistics for the Behavioural Sciences*. New York: McGraw-Hill, 1956.

Swait, J. and J.J. Louviere. " The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *J. Mktg. Res.* 30(1993):305-314.

Swait, J., J.J. Louviere, and M. Williams. "A Sequential Approach to Exploiting the Combined Strengths of SP and RP Data: Application to Freight Shipper Choice." *Transp.* 21(1994): 135-52.

Table 1: List of Attribute Levels and Codes used in Analysis

	Alberta			Saskatchewan		
Attribute (Variable Name/s)	**Survey Levels**	**Codes**		**Survey Levels**	**Codes**	
Access (access1, access2)	No trails, cutlines or seismic lines	-1	-1	Access on foot and or ATV only	-1	-1
	Old trails cutlines or seismic lines, not passable without ATV	-1	-1			
	Newer trails, cutlines or seismic lines, passable with 4WD truck	0	1	Passable with a 4WD vehicle	0	1
	Newer trails, cutlines or seismic lines, passable with 2WD vehicle	1	0	Passable with a 2WD vehicle	1	0
Encounters (enc1, enc2)	No hunters, other than those in my hunting party are encountered	1	0	No other people other than those in your hunting party are encountered	1	0
	Other hunters, hunting on foot are encountered	0	1	Other people, on foot, are encountered	0	1
	Other hunters, on ATV's, are encountered	-1	-1	Other people on ATV's are encountered	-1	-1
	Other hunters, in trucks, are encountered	-1	-1			
Forestry Activity (forest)	No evidence of logging	1		Little or no evidence of logging	1	
	Some evidence of recent logging found in the area	-1		Small (max. width 440m), irregular shaped cutovers, scattered patches of residual tree	-1	
				Large, straight edged clearcut area, no residual trees	-1	
Moose Density (moose1, moose2)	Evidence of less than 1 moose per day	1	0	Evidence of 1 moose every 2 days	1	0
	Evidence of 1 to 2 moose per day	0	1	Evidence of 1 moose per day	0	1
	Evidence of 3 moose per day	-1	-1	Evidence of 3 or more moose every 2 days	-1	-1
	Evidence of more than 4 moose per day	-1	-1			
Road Quality	Mostly paved, some gravel or dirt	1				
	Mostly gravel or dirt, some paved	-1				
Wildlife Species (species1, species2)				Only common species of wildlife	1	0
				Common species of wildlife and 1-2 species you've never seen before	0	1
				Common specie of wildlife, 1-2 species you've never seen before, and a chance of seeing a rare or endangered species	-1	-1
Residence (urban)	Urban (Edmonton)	1		Urban	1	
	Rural (Whitecourt, Hinton, Edson, Drayton Valley)	-1		Rural	-1	

Table 2: Parameter Estimates and Other Information for Ten Canadian Moose Hunting Site Choice Models.

Variables	Saskatchewan data				Alberta data		Joint RP SP data			
	SK-RP1	SK-RP2	SK-SP1	SK-SP2	AB-SP1	AB-SP2	J-SKSP1	J-SKSP2	J-ABSP1	J-ABSP2
SP Intercept [1]			-1.5491	-1.5685	-1.8380	-2.0362	-2.6244	-2.5953	-2.7963	-2.9519
Travel Cost	-1.7226	-2.1550	-0.2499	-0.4565	-0.6479	-0.8680	-0.4541	-0.7788	-0.9906	-1.2545
Access 1	-1.7308	-1.8043	0.2101	0.2097	-0.2278	-0.2623	0.2715	0.2693	-0.3614	-0.3917
Access 2			0.0546	0.0551	0.1795	0.1177	0.1553	0.1493	0.2888	0.1819
Encounters 1	3.0265	3.0353	0.4785	0.4779	0.5262	0.5176	0.7777	0.7602	0.7787	0.7240
Encounters 2			*-0.0381*	-0.0366	-0.0637	-0.0801	-0.1049	-0.0991	*-0.0458*	*-0.0179*
Forestry Activity	*-0.6790* [2]	*-0.6738*	0.1907	0.1904	0.0642	0.0713	0.2547	0.2498	0.1082	0.1231
Moose 1	-1.7190	-1.7571	-0.6205	-0.6240	-1.0406	-1.0364	-1.0549	-1.0386	-1.4998	-1.4111
Moose 2			0.1011	0.1024	0.1829	0.1908	0.2424	0.2391	0.2359	0.2178
Species 1 (SK)	*0.5898*	*0.6400*	-0.1485	-0.1495			-0.2529	-0.2479	*0.0531*	*-0.0280*
Species 2 (SK)			0.0559	0.0564			*-0.0853*	0.0849		
Road Quality (AB)					*-0.0004*	-0.0778			*-0.0055*	-0.1181
Urban Travel Cost (SK)		*0.5216*		0.2557				0.4161		*-0.0262*
Urban Cost Travel (AB)						0.4966				0.7259
Other Urban Interactions?	No	No	No	No	No	Yes	No	No	No	Yes
Relative scale parameter [3]							0.6024	0.6161	0.6699	0.7027
Log-Likelihood	-269.91	-269.58	-3723.7	-3701.8	-3338.5	-3289.2	-4020.4	-3998.1	-3618.6	-3567.3
ρ^2	0.2830	0.2839	0.1693	0.1741	0.2551	0.2662	0.1725	0.1771	0.2552	0.2658
No. of choices	336	336	4080	4080	4080	4080	4416	4416	4416	4416

[1] This intercept represents a dummy variable which equals 1 for the "no hunt option" and 0 otherwise.
[2] Italics indicate lack of significance at the 95% level of confidence.
[3] RP scale normalized to 1.

Table 3: Aggregate Actual and Predicted Trip Distributions and Tests Using RP Holdout Sample Data.

Model	Number of Trips to WMZ[1]											Prediction statistics[2]		
	62	63	64	65	66	67	73	55	68	60	59	$\chi 2$	SAE	Correlation Coefficient
	Predicted Trips													
SK-RP1	10.3	66.4	2.2	11.6	20.9	28.4	2.8	7.6	5.2	35.9	144.7	58.5	119.4	0.9170
SK-RP2	10.7	69.0	2.3	11.3	19.7	27.6	2.8	7.6	5.3	36.5	143.2	56.8	114.7	0.9240
SK-SP1	14.4	30.4	20.3	24.8	26.8	25.2	17.4	22.3	20.6	29.8	103.9	394.7	200.2	0.7588
SK-SP2	15.2	31.7	20.9	24.3	25.8	24.0	16.1	20.8	19.0	30.9	107.3	349.4	191.6	0.7709
AB-SP1	18.4	65.6	15.6	13.8	20.4	34.8	16.2	10.7	8.9	26.6	105.0	155.1	100.6	0.9624
AB-SP2	18.7	64.0	15.4	13.7	19.1	31.6	19.9	10.7	9.3	26.6	107.1	215.7	105.2	0.9521
J-SKSP1	9.2	35.0	12.3	15.1	20.2	21.9	13.1	12.6	11.0	24.2	161.6	182.2	189.9	0.7901
J-SKSP2	10.0	36.5	13.0	14.8	19.0	20.5	12.4	11.7	10.2	25.3	162.6	170.1	188.1	0.7959
J-ABSP1	11.5	71.3	8.2	7.4	15.8	25.7	9.3	5.4	4.2	22.4	154.9	64.9	116.7	0.9262
J-ABSP2	12.9	73.1	11.5	8.2	15.6	29.2	5.8	5.7	4.2	24.1	145.8	40.9	102.1	0.9404
	Actual Trips													
	25.0	101.0	6.0	11.0	18.0	35.0	2.0	4.0	4.0	17.0	113.0			

[1] Total trips=336

[2] These statistics refer to the APS or aggregate $\chi 2$ test, the SAE or sum of absolute errors statistic, and the correlation coefficient computed between the observed and predicted trip distributions.

Table 4: Individual Level Tests of Choice Prediction Success Using RP Holdout Sample Data.

	All Data		
Model	Correlation coefficient	Mean of individual correlation coefficients	σ_n [1]
SK-RP1	0.4959	0.3889	0.0756
SK-RP2	0.5242	0.3927	0.0850
SK-SP1	0.3319	0.3712	0.0579
SK-SP2	0.3486	0.3738	0.0626
AB-SP1	0.5170	0.3806	0.0939
AB-SP2	0.5191	0.3822	0.0967
J-SKSP1	0.3345	0.3760	0.0067
J-SKSP2	0.3513	0.3790	0.0118
J-ABSP1	0.4826	0.3878	0.0442
J-ABSP2	0.5095	0.3880	0.0675

[1] McFadden's normalized prediction success index.
[2] Individuals who took 10 or more trips (4 in total) were removed.

Table 5: Summary of Aggregate and Individual Level Test Results.

	Aggregate Tests			Individual Level Tests		
Rank	$\chi 2$	SAE	Correlation coefficient	Correlation coefficient	Mean of individual correlation coefficients	σ_n
1	J-ABSP2	AB-SP1	AB-SP1	SK-RP2	SK-RP2	AB-SP2
2	SK-RP2	J-ABSP2	AB-SP2	AB-SP2	SK-RP1	AB-SP1
3	SK-RP1	AB-SP2	J-ABSP2	AB-SP1	J-ABSP2	SK-RP2
4	J-ABSP1	SK-RP2	J-ABSP1	J-ABSP2	J-ABSP1	SK-RP1
5	AB-SP1	J-ABSP1	SK-RP2	SK-RP1	AB-SP1	J-ABSP2

Figure 1: Comparison of Actual and Predicted Aggregate Trip Distributions.

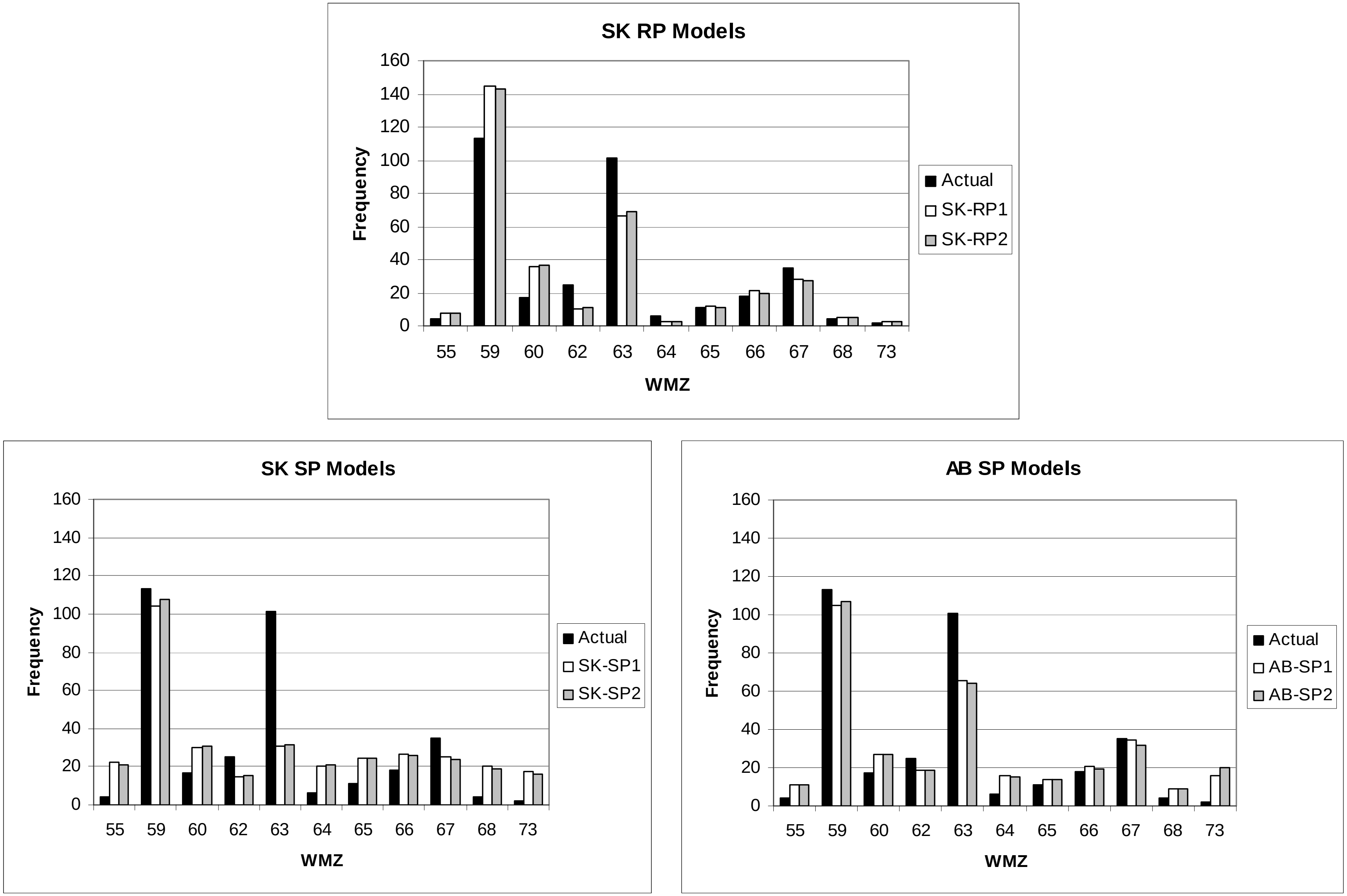

Figure 2: Histograms Comparing Distributions of Individual Correlation Coefficients between Actual and Predicted Trips.