Journal of Choice Modelling, 2(1), pp. 101-117
www.jocm.org.uk

Monte Carlo analysis of SP-off-RP data

Kenneth E. Train[1,*] Wesley W. Wilson[2,†]

[1] Department of Economics, University of California, Berkeley CA 94720-3880
[2] Department of Economics, University of Oregon, Eugene OR

Received 12 September 2008, revised version received 5 December 2008, accepted 11 March 2009

Abstract

SP-off-RP questions are a recent innovation in choice modelling that solicits information from respondents in a different way than standard stated-preference (SP) experiments. In particular, the alternatives and choice of a respondent in a real-world setting are observed, and the respondent is asked whether he/she would choose the same alternative or switch to another alternative if the attributes of the chosen alternative were less desirable in ways specified by the researcher and/or the attributes of non-chosen alternatives were more desirable in specified ways. This construction, called "stated-preference off revealed-preference" (SP-off-RP), is intended to increase the realism of the stated-preference task, relative to standard SP exercises, but creates endogeneity. In this paper, we present a series of Monte Carlo exercises that explore estimation on this type of data, using an estimator that accounts for the endogeneity. The results indicate that, when the variance in the processing error by respondents is the same for SP-off-RP data as for standard SP data, the two solicitation methods provide about the same level of efficiency in estimation, even though the SP-off-RP data contain endogeneity that the estimator must handle while the SP data do not involve endogeneity. For both solicitation methods, efficiency rises, as expected, as the variance of the processing error decreases. These results imply that, if respondents are able to answer SP-off-RP questions more accurately than standard SP questions (and hence have lower variance of processing error), then SP-off-RP data are more efficient that standard SP data. This implication needs to be viewed cautiously, since (i) the actual processing error for each solicitation method is not measured in the current study, and (ii) the results are for the specific data generation processes that are used in the Monte Carlo exercises.

Keywords: Choice modelling, survey methods, stated preference, contingent valuation

* Corresponding author, T: +415-291-1023, F: +415-291-1020, train@econ.berkeley.edu
† T: +541-346-4690, F: +541-346-1243, wwilson@uoregon.edu

ISSN 1755-5345

1 Introduction

Consumers' preferences are often estimated by supplementing data on choices that consumers have made in market settings, called "revealed-preference" (RP) data, with data on choices that consumers say they would make, called "stated-preference" (SP) data. In a typical SP experiment, the researcher constructs hypothetical choice situations, each of which consists of two or more alternatives among which the respondent is asked to choose. The attributes of the alternatives are varied over experiments to provide the variation needed for estimation of underlying preference parameters. The purpose of these SP experiments is to generate variation in attributes when the attributes in the market conditions that produce the RP data exhibit insufficient independent variation to allow precise estimation. Examples include Ben-Akiva and Morikawa (1990), Hensher and Bradley (1993) and Hensher et al. (1999) within a logit specification and Bhat and Castelar (2001) and Brownstone et al. (2000) using mixed logit.

"Pivoting" has been used by some researchers to enhance the realism of SP experiments, by constructing alternatives for SP experiments that are similar to ("pivoted off") an alternative that the agent chose in a market setting. For example, in examining route choice, Rose et al. (2008) asked each respondent to describe a recent trip. Hypothetical routes were designed with times and costs constructed as some percent above or below those of the recent trip. The respondent is then asked to choose among these hypothetical routes. The recent trip with its observed times and cost is either included or excluded from the SP choice set, depending on the design of the experiments. Other applications include Hensher (2004), Caussade et al. (2005), Hensher and Rose (2007), and Greene et al. (2006).

Fowkes and Shinghal (2002) and Train and Wilson (2008) have proposed and implemented an alternative way of constructing SP experiments that has the potential to be more effective in eliciting preferences, while also being more realistic for the respondent than either standard or pivoted SP experiments. The respondent's choice and alternatives in an RP setting is observed and then the respondent is asked which of the RP alternatives he or she would choose if the attributes of the chosen alternative were made worse and/or the attributes of any of the unchosen alternatives were made better. Take, for example, a mode choice situation in which a respondent has chosen bus when car, bus, and rail are available (along with attributes) for the commute to work. The respondent is then asked such questions as: ``Would you have chosen bus if the bus fare were $1.50 instead of $1.00?" or ``Would you have switched to rail if the trains were 10 minutes faster than they are now?" In these questions, the respondent faces the same alternatives as in the RP setting except for a specified change in one or more of the attributes.

A distinguishing feature of these questions is that they incorporate the fact that a change in the respondent's RP choice can occur only if the attributes of the chosen alternative are made worse or the attributes of the non-chosen alternatives are improved. By determining the extent to which the attributes of the chosen alternative must be worsened, or the attributes of non-chosen alternatives improved, in order to induce the respondent to change, the underlying preferences of the respondent are revealed.

Train and Wilson (2008) call this procedure "SP-off-RP" because the stated-preference questions are created from the respondent's revealed-preference choice. While similar to pivoting, the procedure differs from the usual pivoted designs in two important ways. First, with the usual pivoted designs, the respondent faces whatever number of alternatives the researcher constructs and presents to the respondent in the SP task, whereas in SP-off-RP questions the respondent faces the same number of alternatives in the SP task as in the RP task. Second, and related to the first, in SP-off-RP questions, there is a one-to-one correspondence of the SP alternatives to the RP alternatives, whereas in the pivoted experiments cited above each of the SP alternatives corresponds to either one RP alternative (the chosen one) or no specific RP alternative.

SP-off-RP questions provide several potential advantages relative to standard or pivoted SP designs. First, SP-off-RP questions contain a realism that might not be attained by either standard or pivoted SP experiments. This realism occurs because respondents face the same choice situation, with the same alternatives, in the SP-off-RP questions as they faced in the RP setting. The correspondence to their real choice setting can make respondents more able to accurately assess their choices in the SP-off-RP setting. It can also induce respondents to consider the task thoughtfully since the questions are clearly relevant to the respondent’s situation. Second, in standard SP and pivoted SP experiments, the issue necessarily arises of how the respondent assesses or considers the attributes that are not listed in the experiments. For example, in a standard SP experiment for mode choice, the time and cost of the alternatives might be listed, while factors such as risk of delay, the extent of crowding on the bus, whether an easy parking place can be found for the car, etc., are perhaps not included. Inevitably, some attributes are not listed, and it is not clear how the respondent evaluates these non-listed attributes. With SP-off-RP questions, the respondent is asked to consider a change in observed attributes in the RP setting that the respondent faces. The unobserved attributes are by construction, the same as in the RP setting. This commonality of unobserved attributes across the RP and SP-off-RP data can be explicitly represented and tested in the estimation procedure. Third, the task of estimation is to determine respondents’ tradeoffs among attributes as revealed by their choices among alternatives with different attributes. This task is readily served by changing attributes in the directions that are needed to induce a change. Improving an attribute of the chosen alternative cannot change a person's choice and, hence, does not reveal anything about their preferences; neither does worsening the attributes of unchosen alternatives. In standard and the usual pivoted SP experiments, respondents can face choices that reveal little or no information beyond that revealed in their RP choices, since the tradeoffs implied by the RP choice are not taken into consideration in the SP design. In SP-off-RP questions, the attributes are changed in the direction necessary to elicit preference revelation. No matter what the respondent answers in response to these changes, information about preferences is obtained, namely, that the value of the change is either greater than or less than the difference in original utilities.[1]

SP-off-RP questions are also similar to contingent valuation questions, where the respondent is asked how much they would be willing to pay for a specified improvement in attributes. The difference is that the SP-off-RP questions ask the respondents if they would change their choice under specified conditions, and their willingness to pay is inferred through estimation; while in the standard contingent valuation question, the respondents are asked their willingness to pay directly. Since respondents are accustomed to making choices, and seldom need to determine their own maximum willingness to pay, the SP-off-RP design might be expected to solicit more reliable information than the standard contingent valuation question. However, depending on how terms are defined, SP-off-RP questions might be considered a variant of contingent valuation.

[1] This advantage arises only for attributes whose marginal utility has a known sign, such as time and cost of travel for which the marginal utility is negative. If the marginal utility cannot be signed or takes a different sign for different people, then SP-off-RP questions cannot be worded in a way to guarantee a decrease in utility for the chosen alternative or an increase in utility for a non-chosen alternative.

It is most natural to ask RP-off-SP questions in the context of labelled alternatives, such as "car," "bus." and "rail." However, the procedure can also be applied with unlabelled alternatives, as long as the alternatives are specifically identified. For example, the questionnaire might first solicit the respondents' chosen alternative and then ask what other alternatives were available. The SP-off-RP questions would be worded to ask the respondent what they would do if the attributes of their chosen alternative (whatever the respondent said it was) were degraded, or if the attributes of one of the alternatives that the respondent had said was available but not chosen were improved.

The potential advantages SP-off-RP design, however, come at an econometric cost. In particular, as Bradley and Daly (1993) pointed out, the procedure creates endogeneity in the attributes in the SP-off-RP questions, since these attributes are constructed from the respondent's chosen alternative in the RP setting. Unobserved factors in the RP environment affect the respondent's RP choice and, thereby, affect the attributes in the SP-off-RP setting (since these attributes depend on the RP choice.) As discussed above, the unobserved factors in the RP setting carry forward to the SP-off-RP setting. The SP-off-RP attributes are therefore not independent of the unobserved factors, as usually assumed, but rather depend explicitly upon them. This dependence, if ignored, creates inconsistency in the estimator, as Bradley and Daly (1993) described and documented.[2]

Train and Wilson (2008) developed an econometric method that accounts for this endogeneity and provides a consistent and efficient estimator for SP-off-RP data. They applied the method to data from a survey of shippers, using RP data on the shippers' chosen mode and destination, along with SP-off-RP data on whether the shippers' choices would change if the attributes of the chosen mode/destination became worse. In their application, they did not know the true behavioural parameters, and so it was not possible to determine the extent to which the SP-off-RP data provided more precise estimates of them.

In the current paper, we use Monte Carlo methods to examine Train and Wilson's econometric procedure for SP-off-RP data. Since the 'true' parameters are known in Monte Carlo data, we are able to assess the extent to which SP-off-RP data increase efficiency, the bias that arises when the endogeneity in the SP-off-RP data is ignored, and the efficiency of SP-off-RP data relative to standard SP data under different assumptions about the error in each type of data. The findings can be summarised as follows:

- For a sample size of 1000 and the parameters in our base specification, SP-off-RP data reduce standard errors for the relevant parameters by a factor of two relative to RP data alone. This result implies that SP-off-RP data provide as large an efficiency gain as quadrupling sample size (since standard errors are inversely proportional to the square root of sample size.)

- For smaller samples, SP-off-RP data provide an even larger gain in efficiency.

- Ignoring the endogeneity in SP-off-RP data creates significant bias in the estimated parameters.

- The econometric method accounts for the possibility that responses to SP-off-RP questions can be influenced by unobserved factors beyond those than enter the RP choice. These may reflect inattention to the task, the inability to

[2] Endogeneity is not necessarily present for all types of solicitation methods that utilize the observed RP choice of the respondent. Train and Wilson (2008) demonstrate, for example, that endogeneity does not necessarily arise with the pivoted designs described above.

conceptualise the situation, or other quixotic aspects of response. As expected, the efficiency gain from SP-off-RP data rises when the variance of these quixotic errors declines. The same result is obtained, also as expected, for standard SP data.

- When these quixotic errors have the same variance in SP and SP-off-RP data, then the two methods provide about the same degree of efficiency. This result implies that the method that obtains more realistic and less quixotic responses provides the greater efficiency, even after accounting for the potential loss of efficiency that dealing with the endogeneity in SP-off-RP data entails. Since the motivation for using SP-off-RP questions is to enhance the realism of the choice situation, this result implies that SP-off-RP data are more efficient than standard SP data if indeed this motivating concept is correct.

There are several potential limitations of SP-off-RP designs. We do not address these limitations in our Monte Carlo analyses, leaving them for future investigation. First, respondents might exhibit inertia, by which they say they would remain with their chosen alternative in the face of changes that would, if actually experienced, induce them to switch. Of course, there is inertia in actual choices, and it is possible that the opposite direction of bias occurs: that the ease of saying "I'd switch" in response to a survey question understates the inertia, or switching costs, that actually arise. Second, respondents might not answer truthfully, but instead answer in ways that they think will affect the outcome that they believe is being investigated. Train and Wilson (2008) discuss the issue of "incentive compatibility" in relation to SP-off-RP questions, discussing conditions under which these questions can be expected to elicit truthful answers. For the purposes of our Monte Carlo simulations, we assume that respondents behave as specified by the model, answering truthfully. Third, the procedure requires that the researcher obtain information about the RP alternatives. In contrast, standard SP experiments can be administered and used in estimation without any RP data. While it is customary to combine SP with RP data, it is not necessary; whereas the use of SP-off-RP questions necessitates the collection of RP data. Fourth, the difference between measured versus perceived attributes can take particular importance in SP-off-RP questions, depending on how the data for the RP alternatives are obtained. The researcher might ask the respondent to provide information on the attributes of the alternatives in the RP setting, in which case the variables entered by the researcher are those perceived by the respondent. Alternatively, the researcher might measure the attributes "objectively", in which case they can differ from the respondent's perceptions. While this issue arises in all choice modelling, it has a new implication in the context of SP-off-RP questions. In particular: Any difference between perceived and measured attributes constitutes part of the unobserved component of utility. When the measured attributes are changed in the SP-off-RP questions, the perceived attributes need not change the same amount, such that the difference between measured and perceived also changes. In this case, the unobserved component is not the same in the SP-off-RP setting as in the original RP setting, and the modelling strategy in this paper would need to be modified to account for this difference.

In the following section, we describe the econometric method for estimating parameters using SP-off-RP data. In section 3, we describe the specification of the Monte Carlo experiments and their results.

2 Econometrics of SP-off-RP data

We describe a fixed coefficient specification first and then generalise to random coefficients.

2.1 Fixed coefficients

Each agent faces a choice among discrete alternatives in an RP setting. As is common, the utility that agent n obtains from alternative j is denoted U_{jn}, which is decomposed into observed and unobserved components:

$$U_{jn} = \beta x_{jn} + \varepsilon_{jn} \tag{1}$$

We assume that ε_{jn} is iid type one extreme value, with the result that the model of the RP choice is a standard logit and can be estimated with conventional methods.

The SP-off-RP data are constructed from the RP response. To obtain the SP-off-RP data, the researcher gives the agent a series of choice tasks in which the attributes of the alternatives in the RP setting are changed based on the agent's choice in the RP setting, making the attributes of the chosen alternative worse and/or the attributes of the non-chosen alternatives better. The researcher constructs T choice tasks, each consisting of the same alternatives as in the RP setting but with changed attribute levels. Let $\tilde{x}^i_{jnt}$ denote the attributes for alternative j in choice task t based on alternative i having been chosen in the RP setting.[3] The utility of each alternative in these choice tasks is assumed to take the form:

$$W_{jnt} = \beta \tilde{x}^i_{jnt} + \varepsilon_{jn} + \mu^*_{njt} \tag{2}$$

where μ^*_{njt} is a new error term. Specifically, under this specification, the agent makes an assessment of the alternatives using the same coefficients β and same ε_{jn} as in the RP setting, but the attribute of the RP choice is made worse or an alternative made better. In addition, there is also an additional error term (μ^*_{njt}) that reflects, e.g., inattention by the agent, pure randomness in the agent's responses, or other quixotic aspects of the choice task. Importantly, the unobserved factors ε_{jn} that affect the agent's choice in the RP setting carry forward to the choice task, since these unobserved factors are not changed (the assumption that the same β and ε_{jn} enter the RP and SP-off-RP choices can be tested, but for our discussion we take the specification as given). Let the new error μ^*_{njt} be iid extreme

[3] An important issue for future research is the efficient design of the changes, i.e., of $\tilde{x}^i_{jnt}$. For our Monte Carlo simulations, and in the applications in Train and Wilson (2008), changes in attributes are selected randomly from a specified range. With standard SP experiments, various designs have been found to provide considerably greater efficiency than random selection of attributes, at least for small to moderately sized samples. See, e.g., Rose and Bleimer (2008). It can be expected that similar improvements are potentially available from more efficient design of SP-off-RP questions.

value with scale 1/λ. A large value of parameter λ indicates that there are few quixotic aspects to the SP-off-RP choices, and the agent chooses essentially the same as in an RP situation under the new attributes. Utility can be equivalently expressed as:

$$W_{jnt} = \lambda\beta\, \tilde{x}^{i}_{jnt} + \lambda\varepsilon_{jn} + \mu_{njt} \tag{3}$$

where now μ_{njt} is iid extreme value with unit scale. The SP-off-RP choices are, therefore, standard logits with ε_{jn} as an extra explanatory variable. Since the ε_{jn}'s are not observed, these logits must be integrated over the conditional distribution of these RP errors. In particular, the probability of alternative *k* being chosen in choice task *t* given that the agent chose alternative *i* in the RP setting is

$$P^{i}_{knt} = \int \frac{e^{V_{knt}(\varepsilon_{kn})}}{\sum e^{V_{jnt}(\varepsilon_{jn})}}\, f(\varepsilon_n \,|\, U_{in} > U_{jn}\ \forall j \neq i)\, d\varepsilon_n \tag{4}$$

where $V_{jnt}(\varepsilon_{jn}) = \lambda\beta\, \tilde{x}^{i}_{jnt} + \lambda\varepsilon_{jn}$ and *f* is the density of $\varepsilon_n = \langle \varepsilon_{1n},...,\varepsilon_{Jn} \rangle$ conditional on alternative *i* having been chosen in the RP setting. This choice probability is a mixed logit, with mixing over ε_n. It is simulated by taking draws from *f*, calculating the logit formula for each draw, and averaging the results. Train and Wilson (2008) derive the conditional density of ε_n based on earlier work by Anas and Feng (1988) and show how to take draws from it.

Under the assumption that μ_{njt} is independent over choice tasks, the probability of the agent's choices in all *T* tasks is the product of logits for the *T* choices, integrated over the conditional distribution of ε_{jn}. The probability of the RP and SP-off-RP choices, which enters maximum likelihood estimation, is the product of (i) the logit probability of the RP choice and (ii) the mixed logit probability of the sequence of SP-off-RP choices conditional on the RP choice:

$$P_n = \int \prod_t \left[\frac{e^{V_{k_t nt}(\varepsilon_{k_t n})}}{\sum e^{V_{jnt}(\varepsilon_{jn})}} \right] f(\varepsilon_n \,|\, U_{in} > U_{jn}\ \forall j \neq i)\, d\varepsilon_n \cdot \frac{e^{\beta x_{in}}}{\sum e^{\beta x_{jn}}} \tag{5}$$

The assumption that μ_{njt} is independent over choice tasks, while maintained in the current analysis, is perhaps overly restrictive, since it implies that the quixotic issues affecting the agent take a new form with each choice task. The assumption can be relaxed with a corresponding change in the probability formula. For example, μ_{njt} might take an error-components form, consisting of an independent extreme value part and a part that is the same for a given agent over choice tasks. With this specification, the constant part becomes a new term added to *V* in equation (5), with integration over the distribution of this new term as well as the integration over ε_n.

2.2 Random coefficients

Utility is the same as above except that β is now random with density *h(β)* with underlying parameters (not given in the notation) denoting, e.g., the mean and variance of β. The probabilities are the same as above, except the formulas are now mixed over the distribution of β. The probability that enters the likelihood function is $PR_n = \int P_n(\beta)\, h(\beta)\, d\beta$ where $P_n(\beta)$ is given by equation (5) with β treated as an argument.

3. Monte Carlo analysis

To explore the properties of estimation with SP-off-RP data, we start with a specification that consists of two alternatives, labelled 1 and 2, with two explanatory variables, labelled *x* and *z*. This can easily be adapted to multiple alternatives and additional variables. Utility contains an alternative-specific constant (α_i) , one variable (z_{in}) with a fixed coefficient, and the other variable (x_{in}) with a random coefficient:

$$U_{in} = \alpha_i + \theta \cdot z_{in} + \beta_n \cdot x_{in} + \varepsilon_{in}, \quad i = 1, 2, \quad n = 1, \ldots, N \tag{6}$$

with

ε_{in} ~ *iid extreme value,*

$\beta_n \sim N(\bar{\beta}, \sigma^2)$

The true parameters are specified to be:

$\alpha_1 = 1, \quad \alpha_2 = 0, \; \theta = 1, \quad \bar{\beta} = 1, \quad \sigma = 0.5$.

Each variable for each alternative is specified to be distributed uniformly between 2 and 4, such that the difference between the two alternatives ranges from -2 to +2 for each of the two variables (in the sections below, each of these elements of the data generation process is revised to examine their impact on the estimator).

The agent chooses alternative 1 iff $U_{1n} > U_{2n}$ and otherwise chooses alternative 2. Define $d_n^{\,1} = 1$ if agent *n* chooses alternative 1, = 0 otherwise; and define $d_n^{\,2}$ similarly. This choice, and the value of the variables *x* and *z*, are the RP data. Thus, again, we observed the choice set and attributes along with the attributes in the RP data. We now specify the SP-off-RP data. Only one choice task is given to each agent. We leave a multiplicity of added choice tasks to future research. If alternative *i* is chosen in the RP setting, the value of x_{in} is lowered by r_n proportion, where r_n is uniformly distributed between 0 and 1. Utility in the SP-off-RP situation becomes

$$W_{in} = \lambda(\alpha_i + \theta \cdot z_{in} + \beta_n \cdot (x_{in} - r_n d_n^{\,i} x_{in}) + \varepsilon_{in}) + \mu_{in}, \quad i = 1, 2 \quad n = 1, \ldots, N \tag{7}$$

where the subscript *t* is omitted. The new error μ_{in} is specified to be iid extreme value with unit scale after standardising for the true scale, which is specified to be $\lambda = 4$. This value of the scale was chosen for our initial specification because it is similar to that estimated by Train and Wilson (2008). The agent chooses alternative 1 iff $W_{1n}>W_{2n}$ and otherwise chooses alternative 2. A central point is that the attribute level $(x_{in}-r_n d_n{}^i x_{in})$ is correlated with ε_{in} since the agent's RP choice, $d_n{}^i$, depends on ε_{in}. It is this correlation that constitutes the endogeneity that arises in SP-off-RP data.

Each sample consists of 1000 agents. The sample data are simulated and the parameters are estimated 100 times. In each estimation, 100 randomised Halton draws are used to simulate the integral over the random coefficient β_n and the conditional errors ε_{1n} and ε_{2n} in W_{1n} and W_{2n}.

The results are summarised in the top part of Table 1 (the bottom part of the Table contains estimates on the RP data alone, which we provide for comparison and discuss later). The mean estimates are very close to the true parameters, with none of the differences being statistically significant.[4] Also, the standard deviations of the estimates are very similar to the mean standard errors, which imply that the standard errors provide reliable information, on average, about the expected sampling error in the point estimates. With the exception of the scale parameter, the standard deviations of the standard errors are quite small, indicating that the standard errors for any one sample (i.e., in any one run) are useful indications of the expected sampling error in the point estimates.

The central point of this research is whether, and the extent to which, the SP-off-RP data provide better estimates than the RP data alone. The bottom part of Table 1 gives the results of estimation on the RP data alone (i.e., the agent's choice between alternatives 1 and 2 based on U_{1n} and U_{2n}) without the SP-off-RP data. The mean estimates are close to the true values, though for two of the parameters (the fixed coefficient and the mean of the random coefficient) the hypothesis that the mean equals the true value can be rejected at the 95% level. It is noteworthy that, in all cases, the standard deviations of the estimates are larger using only the RP data than when using the combined RP and SP-off-RP data. Not surprisingly, since the SP-off-RP exercise changes the value of *x*, which has the random coefficient, and, the greatest effect is observed in the estimated parameters of the random coefficient. The use of the SP-off-RP data (Table 1) reduces the standard deviation of the estimates obtained from the use of RP data (Table 2) by over half, from 0.1361 to 0.0586 for the mean of the random coefficient and from 0.3821 to 0.1434 for the estimated standard deviation. To put this improvement in perspective, using the SP-off-RP data is equivalent to more than quadrupling sample size with RP data alone[5] (since a four-fold increase in sample size reduces asymptotic standard errors by two.) Interestingly, the estimates of the fixed coefficient and the alternative-specific constant are also improved by the SP-off-RP data, even though the SP-off-RP exercise only changed the variable with the random coefficient. The standard deviations of these parameters drop by more than 20% when using the SP-off-RP data. The SP-off-RP data allow more precise estimation of these parameters because an agent's response to a change in one variable (i.e., the one changed in the SP-off-RP exercise) depends on the difference in the utility between alternatives prior to the change and, therefore, reveals information about all utility parameters.

[4] Since there are 100 runs, the standard deviation of the mean is one-tenth the standard deviation of the point estimates. The t-statistic for the hypothesis that the mean of the sampling distribution of point estimates of, for example, the intercept is equal to 1.0 (it's true value) is (1-0.9940)/(0.0753/10) = 0.80.

[5] Given the cost of sampling, the result suggests significant savings of approximately 75 percent.

Table 1: Monte Carlo Results for Basic Specification

	Alternative-specific constant	**Fixed coefficient**	**Random coefficient: mean**	**Random coefficient: standard dev.**	**Scale**
True value	1	1	1	0.5	4
Model on RP and SP-off-RP data					
Point estimates:					
Mean	0.994	0.9988	0.9984	0.4843	4.374
Standard deviation	0.0753	0.0961	0.0586	0.1434	2.545
Standard errors:					
Mean	0.0769	0.095	0.0639	0.1383	2.723
Standard deviation	0.0038	0.0068	0.0108	0.0314	4.747
Model on RP data only					
Point estimates:					
Mean	1.0098	1.0278	1.0502	0.5473	-
Standard deviation	0.0988	0.1208	0.1361	0.3821	-
Standard errors:					
Mean	0.0988	0.1207	0.1496	0.5691	-
Standard deviation	0.0139	0.0147	0.0349	0.1666	-

The benefits of SP-off-RP data are described above, but, as noted earlier, their use comes at a cost. Specifically, the use of SP-off-RP data necessitates the need to model endogeneity. To examine the effect of ignoring the endogeneity, estimation was performed using the SP-off-RP data i.e., containing both the RP data and the SP data constructed from the RP response, without controls for the endogeneity of the attributes of the SP component. This procedure is denoted estimation on "SP" data, in quotes. Specifically, the RP and "SP" data were combined for joint estimation, and a separate scale was allowed for the "SP" choice, as is customary when combining RP and SP data. The results are summarised in Table 2. The primary point of Table 2 is that mean estimates are *all* significantly different from the true values.[6] The result indicates that estimation on SP-off-RP data as if they were standard data can cause substantial estimation error, and points to the need to model the endogeneity.

Table 3 summarises results with estimation on 250 observations instead of 1000. The top part of the panel gives results for estimation on the RP and SP-off-RP data, while the bottom part has results for estimation on the RP data alone. Since sample size is reduced by four, the standard deviations of the estimates and the mean standard errors are expected to double, provided that the smaller sample size is still sufficiently large for the asymptotic distributions to be approximately accurate. As can be seen in the top part of the table, the mean estimates are close to the true values, with no significant differences even with the smaller sample size. The standard deviations are similar to the mean standard errors, and both are about twice as large as their values in Table 1 with 1000 observations.

[6] The differences are most prominent in the standard deviation of the random coefficient (whose mean estimate is more than twice the true value) and the scale parameter (whose mean estimate is less than a quarter of the true value.) It is not clear why the error is more concentrated in these parameters than the others, and we have not investigated whether the pattern arises under other specifications.

Table 2: Monte Carlo Results for Basic Specification, ignoring endogeneity

	Alternative-specific constant	**Fixed coefficient**	**Random coefficient: mean**	**Random coefficient: standard dev.**	**Scale**
True value	1	1	1	0.5	4
Point estimates:					
Mean	1.1288	1.1513	0.9555	1.2445	0.638
Standard deviation	0.0942	0.1236	0.1142	0.2026	0.089
Standard errors:					
Mean	0.0934	0.1131	0.1184	0.1982	0.097
Standard deviation	0.0061	0.0084	0.0143	0.0206	0.013

These results imply that the asymptotical distribution still seems to serve as a good approximation with as few as 250 observations.

The results on the RP data alone indicate that the SP-off-RP data are even more useful for small samples than for the larger sample. First, the standard deviation of the estimates based on RP data alone increase by considerably more than twice when reducing sample size from 1000 to 250 (next to last row of Table 3 compared with third from last row of Table 1). Second, even with these larger standard deviations, the mean estimates are significantly different from their true values for three out of the four parameters (third from last row of Table 3). These two results indicate that the sample size is too small for the asymptotic properties to be exhibited when estimation is performed on the RP data alone. The inclusion of the SP-off-RP data reduces by a factor of over three the standard deviations of the estimates of the parameters of the random coefficients (top part of Table 3 compared with bottom part). This improvement is greater than the two-fold improvement that was obtained with a sample of 1000, discussed above.

We next examine various aspects of the specification to assess the impact of each element on the efficiency of the estimator. In particular, we make each of the following changes in specification:

a. Reduce the range of the explanatory variables to be uniform between 2.5 and 3.5 instead of 2 and 4, such that the difference between alternatives ranges from -1 to +1 instead of -2 to +2.

b. Reduce the level of the explanatory variables to be uniform between 1 and 3 instead of 2 and 4. The difference between alternatives still ranges from -2 to +2. The reduction in level changes the magnitude of the reduction in x for the chosen alternative in the SP-off-RP data. (Since x is reduced by a proportion of its value, the reduction is smaller in magnitude when the level of x is smaller.)

c. Reduce the range of reductions in x, such that the proportion reduction r_n is uniformly distributed between 0.25 and 0.75 instead of 0 to 1.

d. Reduced the scale from 4 to 2, thereby doubling the standard deviation of the error associated the SP-off-RP choice.

e. Reduce the scale even further to 0.5, thereby increasing the standard deviation of the processing error by a factor of eight relative to the original specification and by a factor of four relative to the specification in (d).

Table 3: Monte Carlo Results for Basic Specification, 250 observations

	Alternative-specific constant	**Fixed coefficient**	**Random coefficient: mean**	**Random coefficient: standard dev.**	**Scale**
True value	1	1	1	0.5	4
Model on RP and SP-off-RP data					
Mean estimates	1.0073	1.0054	1.0220	0.4846	4.6784
Std dev. estimates	0.1399	0.1741	0.1252	0.2784	6.4835
Mean S.E.s	0.1583	0.1911	0.1250	0.2812	5.8497
Model on RP data only					
Mean estimates	1.0813	1.0800	1.1854	0.8262	
Std dev. estimates	0.2460	0.3013	0.4646	0.9943	
Mean S.E.s	0.2185	0.2638	0.3523	1.0448	

Each of these changes is designed to decrease the efficiency of the estimator by decreasing either the variation in the data (specifications a-c) or the precision of the agents' responses to the SP-off-RP question (specifications d and e). Table 4 summarises the results. The mean estimates are given in the top part of the table and the standard deviations in the bottom. For comparison, the first row of each part gives results for the original specification (i.e., repeats the information from Table 1).

The means are close to the true values in all specifications. Using a t-test at the 95% confidence level, the hypothesis that the mean is equal to the true value is rejected in only four instances, whose means are given in bold in the table. Since there are a total of 30 such tests, the expected number of rejections when the hypothesis is true is 1.5, and the probability of obtaining 4 or more rejections is 0.06. The hypothesis that all the means are equal to their true values can, therefore, be rejected at the 95% level but not that the 97% level. In any case, the differences are small and the significant ones are not concentrated in any one specification.

Specification (a) reduces the range of the explanatory variables relative to the base specification. As expected, the standard error of the parameters associated with both variables, as well as the scale parameter, rise relative to those in the base specification. In specification (b), the level of x and z for each alternative decreases by 1. This change does not affect the difference in variables between alternatives in the RP choice, since the reduction is applied to each alternative. However, in the SP-off-RP data, the value of x for the chosen alternative is reduced by a proportion, while the value of x for the non-chosen alternative is not changed. The effect of the new specification, therefore, is to reduce the range of x in the SP-off-RP question. As expected, the standard deviations of the parameters for the coefficient of x and the scale of the SP-off-RP error rise. Specification (c) also decreases the range of x in the SP-off-RP data, by decreasing the range of the proportion by which x for the chosen alternative is reduced. As with specification (b), the standard deviations of the parameters of the coefficient of x and the scale parameter rise. Specification (d) and (e) increase the standard deviation of the additional error that enters the SP-off-RP choices, which, intuitively, makes these choices more "noisy" and, hence, less useful for estimation of the true behavioural parameters. The scale is estimated fairly

Table 4: Monte Carlo Results for Variations on Basic Specification

	Alternative-specific constant	**Fixed coefficient**	**Random coefficient: mean**	**Random coefficient: standard dev.**	**Scale**
Mean estimates					
Base specification	0.9940	0.9988	0.9984	0.4843	4.3736
a. X and Z uniform 2.5-3.5	0.9977	1.0071	0.9914	**0.4509**	4.7696
b. X and Z uniform 1-3	0.9948	0.9877	**0.9853**	0.4740	4.8225
c. R_n uniform .25-.75	1.0139	**1.0272**	1.0058	0.4767	4.3213
d. Scale =2	0.9915	**0.9805**	0.9912	0.4672	2.0782
e. Scale = 0.5	0.9902	0.9813	0.9869	0.4455	0.4966
Standard deviations					
Base specification	0.0753	0.0961	0.0586	0.1434	2.5449
a. X and Z uniform 2.5-3.5	0.0712	0.1480	0.0713	0.1849	4.1215
b. X and Z uniform 1-3	0.0739	0.0880	0.0675	0.1854	4.7024
c. R_n uniform .25-.75	0.0719	0.1058	0.0725	0.1609	5.3613
d. Scale =2	0.0770	0.0912	0.0703	0.1821	0.6913
e. Scale =0.5	0.0851	0.1050	0.0875	0.3702	0.0910

precisely in each case: the mean estimate is 2.0782 when scale is 2.0, and 0.4966 when scale is 0.50. The standard deviations rise, as expected, but far less than the increase in the standard deviation of the error. For example, the standard deviation of the estimates of the standard deviation of the coefficient of x (which is the parameter that is most affected by the change in scale) rises from 0.14 to 0.18 when the standard deviation of the error doubles, and rises from 0.14 to 0.37 when the standard deviation of the error rises by a factor of eight.

An important issue is whether, or the conditions under which, SP-off-RP data provide more information for estimation than standard SP experiments. To address this issue, simulations were performed with each agent presented with a standard SP experiment rather than an SP-off-RP experiment. For the first comparison, each agent is given a choice between two alternatives that differ in x and the identity of the alternative (which determines the alternative-specific constant.) This set-up, with only x varying, corresponds to the SP-off-RP choice in which the value of x was changed.[7] Specifications with both x and z varying are considered below. Utility in the SP choice is assumed to take the same form as in the RP choice, with each agent using the same parameters as in their RP choice, except that the standard deviation of the error in the SP choice differs from that in the RP

[7] The values for x_{in} in the SP experiments were generated in the same way as for the RP data, by randomly drawing a value from a uniform distribution between 2 and 4, such that the difference in x between the two alternatives ranged from -2 to +2. For the SP experiments with z, each z_{in} is generated similarly.

choice by a factor (1/ λ). The model was estimated on the combined RP and SP data, with a separate scale λ for the SP data and all other parameters being the same. Estimation of a separate scale for RP and SP data when combining the two is standard practice; see, e.g., Ben-Akiva and Morikawa (1990), Hensher and Bradley (1993), Louviere *et al*. (2000), and Train (2003, section 7.2). It is also standard practice to estimate separate alternative-specific constants on the RP and SP data. We instead use the same constant for both types of data (both in simulation of the choices and in estimation), which increases the efficiency of the SP estimator in our analysis. The standard deviations of the parameter estimates on the RP/SP data in our analysis are, therefore, smaller than would be expected under standard practice.[8]

The results are summarised in Table 5, which, for comparison, also contains results for estimation using the SP-off-RP data (repeated from previous tables.) Estimation is performed with true scale set at 4, 2 and 0.5, with smaller scale indicating greater processing error in the SP choices (i.e., larger standard deviation of the unobserved portion of utility in the SP choices). The scale parameter for the SP data is not exactly comparable to the scale parameter for the SP-off-RP data. For the SP data, the scale reflects the standard deviation of the unobserved portion of utility in the SP choice relative to that in the RP choice. For the SP-off-RP data, the scale reflects the standard deviation of the *extra* error that is added to the unobserved portion of utility in the RP setting. The same value of the scale parameter, therefore, implies larger total error in the SP-off-RP utility than in the SP utility. This difference in the meaning of the scale parameter implies that the comparisons in Table 5 are biased in favour of the SP data over the SP-off-RP data, since the set-up gives a larger error for the SP-off-RP data than the SP data.

The mean estimates based on combined RP and SP data are similar to the true values, with the hypothesis of equality to the true value being rejected only three times in the 20 tests (shown in bold). In this regard, the SP data perform about the same as the SP-off-RP data, which obtained two rejections out of the 20. For a given level of the scale parameter, the standard deviations of the estimates using SP data are similar to those using SP-off-RP data, with some being smaller and some larger. As the scale drops (i.e., as the "processing" error attached to SP questions becomes greater), the standard deviations of the estimates rise under both approaches. These two results combined imply that the procedure that has the lower processing error can be expected to provide more precise estimates. One of the motivations for the use of SP-off-RP questions instead of SP experiments is that, by asking questions in relation to the respondent's a real-world choice, the respondent is more able to meaningfully assess the hypothetical situation. If this conjecture is true, or, more precisely, if the processing error in SP-off-RP choices is, indeed, less than in SP experiments, then these simulation results indicate that greater estimation efficiency is obtained with SP-off-RP data than with SP data.

[8] The constants can be allowed to differ between the RP and either the SP or the SP-off-RP choices, to reflect the possibility that the average of unobserved factors (which the constants capture) is different in the two settings. We have not investigated the implications of this generalization, but note that a finding that constants differ is perhaps more problematic with SP-off-RP questions, which are motivated by the notion that the unobserved portion of utility carries from the RP setting to the SP-off-RP setting, than with standard SP experiments, where no presumption is made that the unobserved terms being the same in the SP and RP settings.

Table 5: Monte Carlo Results for SP Data and SP-off-RP Data

	Alternative -specific constant	**Fixed coefficient**	**Random coefficient: mean**	**Random coefficient: standard dev.**	**Scale**
Mean estimates					
Scale = 4					
SP-off-RP	0.9940	0.9988	0.9984	0.4843	4.3736
SP	1.0058	1.0049	1.0005	0.4920	**4.1270**
Scale = 2					
SP-off-RP	0.9915	**0.9805**	0.9912	0.4672	2.0782
SP	1.0061	1.0050	1.0001	0.4691	2.0130
Scale = 0.5					
SP-off-RP	0.9902	0.9813	0.9869	0.4455	0.4966
SP	1.0081	1.0107	1.0082	**0.5809**	**0.4850**
x and z vary					
SP-off-RP	0.9965	1.0039	0.9997	0.4884	**3.8364**
SP	1.0058	1.0066	0.9932	0.4895	4.0210
Standard deviations					
Scale = 4					
SP-off-RP	0.0753	0.0961	0.0586	0.1434	2.5449
SP	0.0819	0.1063	0.0974	0.1305	0.5883
Scale = 2					
SP-off-RP	0.0770	0.0912	0.0703	0.1821	0.6913
SP	0.0848	0.1073	0.1002	0.1950	0.2079
Scale = 0.5					
SP-off-RP	0.0851	0.1050	0.0875	0.3702	0.0910
SP	0.0939	0.1138	0.1260	0.3473	0.0685
x and z vary					
SP-off-RP	0.0735	0.0442	0.0486	0.0743	0.5672
SP	0.0691	0.0706	0.0756	0.0677	0.4038

Table 5 contains one last comparison. In the specifications considered so far, only x was varied in the SP-off-RP and SP data. It is, of course, customary to include a series of SP-off-RP or SP tasks with each relevant variable varying. We next consider, therefore, SP-off-RP questions about both *z* and *x*, and SP experiments that contain both variables. The specification is the same as the base specification, with scale parameter of 4. For the SP-off-RP data, each agent is asked two questions: one question (the same as in earlier specifications) about how they would respond if *x* for their chosen alternative were reduced by a certain proportion, and a second question, that is similar but for a reduction in *z* for their chosen alternative. The outcome consists of the agent's choice between the original RP alternatives, their choice when x for their chosen RP alternative is reduced, and their choice when *z* for their chosen RP alternative is reduced. SP data are specified analogously. Two SP experiments are administered for each agent, with *x* and *z* varying over alternatives and experiments. The outcome consists of the agent's choice between the RP alternatives and their choices in the two SP experiments.

The last rows in both parts of Table 5 summarise the results for these specifications. The standard deviations are considerably lower than with only *x* varying. For example, the standard deviation of the fixed coefficient of *z* drops from 0.0961 when asking an SP-off-

RP question about *x* only to 0.0442 when asking questions about both *x* and *z*. Similarly, for SP experiments, the standard deviation drops from 0.1063 using one experiment with x varying to 0.0706 using two experiments with both *x* and *z* varying. The standard deviations are about the same for the two methods when both *x* and *z* are varied, with the SP-off-RP data obtaining a lower standard deviation for some parameters (viz., the fixed coefficient and the mean of the random coefficient) and the SP data obtaining smaller standard deviations for the other parameters (the intercept, scale, and standard deviation of the random coefficient). These results confirm the earlier statement based on one SP-off-RP and SP task that, when the processing error for the two types of data are the same, SP-off-RP questions and SP experiments provide about the same level of estimation efficiency. The researcher's decision of which method to use depends largely, therefore, on which method the researcher expects will induce less processing error by the respondents.

5. Conclusions

SP-off-RP data are generated by changing the attributes of alternatives in an RP setting on the basis of the agent's choice in that setting. The primary advantages of such data are that: 1) as with any form of SP tasks, the data can contain substantially more variation in the attributes underlying the choice than is commonly observed in RP data; and 2) the SP-off-RP data are constructed from the revealed choice made by the agent and, as such, overcome the common criticism of SP data i.e., the lack of realism. Yet, since the SP-off-RP data are endogenous, estimation is more complicated. The Monte Carlo results presented in this paper suggest that the added complication is repaid in potentially substantial gains in efficiency. In our base specification, models estimated with SP-off-RP data obtained approximately the same level of efficiency as models estimated with RP data, but with only about ¼ of the observations. This approach, therefore, can provide substantial savings in sampling costs.

Responses to the SP-off-RP questions may differ for a variety of unobserved factors that are unrelated to the RP error; e.g., the respondent may not be attentive to the task or may tend to answer randomly. Such issues also arise in standard SP experiments. The estimator for SP-off-RP designs explicitly allows for quixotic responses. As expected, the efficiency gain from SP-off-RP data rises when this error variance declines. Importantly, our Monte Carlo results indicate that SP-off-RP data provide greater efficiency than standard SP data if, as expected, the variance of this response error is lower in SP-off-RP data than in SP data.

Finally, the Monte Carlo experiments suggest that it is critically important to model the endogeneity in SP-off-RP data. Indeed, if one uses SP-off-RP data to estimate the parameters of a choice model, but ignores the endogenous construction of the data, significant bias can be introduced. Thus, for a given sample size, there are potential efficiency gains from an SP-off-RP design, but these gains can only be attained when the estimation procedure appropriately reflects the endogeneity created by SP-off-RP questions.

Acknowledgements

The authors gratefully acknowledge research support from the Navigation and Economics Technology Program of the Institute for Water Resources.

References

Anas, A. and Feng, C., 1988. Invariance of expected utilities in logit models. Economic Letters, 27, 57-78.

Ben-Akiva, M. and Morikawa, T., 1990. Estimation of switching models from revealed preference and stated intentions. Transportation Research Part A, 24, 485-495.

Bhat, C. and Castelar, S., 2001. A unitfied mixed logit framework for modeling revealed and stated preferences: Formulation and application to congestion pricing in the San Francisco bay area. Transportation Research Part B, 36, 577-669.

Bradley, M., and Daly, A., 1993. New analysis issues in stated preference research. In Proceedings of Seminar-D. PTRC 21st Summer Annual Meetings, London, reprinted in J. de D. Ortuzar, ed., Stated Preference Modeling Techniques, PTRC Education and Research Services, London, 75–89.

Brownstone, D., Bunch, D. and Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fueled vehicles. Transportation Research Part B, 34, 315-338.

Caussade, S., Ortuzar, J. de D., Rizzi, L. and Hensher, D., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B, 39, 621-640.

Fowkes, A. and Shinghal, N., 2002. The Leeds adaptive stated preference methodology. In R. Danielis, ed., Freight Transportation Demand and Stated Preference Experiments, FrancoAngeli, Milan.

Greene, W., Hensher, D., and Rose, J., 2006. Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. Transportation Research Part B, 40, 75-92.

Hensher, D., 2004. Accounting for stated choice design dimensionality in willingness to pay for travel time savings. Transportation Research Part B, 38, 425-446.

Hensher, D. and Bradley, M., 1993. Using stated response data to enrich revealed preference discrete choice models. Marketing Letters, 4, 39-152.

Hensher, D., Louviere, J. and Swait, J., 1999. Combining sources of preference data. Journal of Econometrics, 89, 197-221.

Hensher, D. and Rose, J., 2007. Development of commuter and noncommuter mode choice models for the assessment of new public transport infrastructure projects: A case study. Transportation Research Part A, 41, 428-443.

Louviere, J., Hensher, D. and Swait, J., 2000. Stated Preference Methods: Analysis and Applications, Cambridge University Press, New York.

Rose, J. and Bliemer, M., 2008. Stated preference experimental design strategies. In Handbook of Transport Modelling, ed. D.A Hensher and K.J. Button, Elsevier Oxford, United Kingdom, 151-79.

Rose, J., Bliemer, M., Hensher, D. and Collins, A., 2008. Designing efficient stated choice experiments in presence of reference alternatives. Transportation Research Part B, 42, 395-406.

Train, K., 2003. Discrete Choice Methods with Simulation, Cambridge University Press, New York.

Train, K. and Wilson, W., 2008 Estimation on stated preference experiments constructed from revealed-preference choices. *forthcoming*, Transportation Research Part B, 42, 191-203.