A control-function approach to correct for endogeneity in discrete choice models estimated on SP-off-RP data and contrasts with an earlier FIML approach by Train & Wilson

C. Angelo GUEVARA

Departamento de Ingeniería Civil,
Universidad de Chile,
crguevar@ing.uchile.cl,
+56229784380,
Blanco Encalada 2002, Santiago, Chile

Stephane HESS

Institute for Transport Studies & Choice Modelling Centre,
University of Leeds,
s.hess@its.leeds.ac.uk,
+44 113 343 6611, 36-40
University Road, LS2 9JT, Leeds

C. A. Guevara and S. Hess

ABSTRACT

It is common practice to build Stated Preference (SP) attributes and alternatives from observed Revealed Preference (RP) choices with a view to increasing realism. While many surveys pivot all alternatives around an observed choice, others use more adaptive approaches in which changes are made depending on what alternative was chosen in the RP setting. For example, in SP-off-RP data, the alternative chosen in the RP setting is worsened in the SP setting and other alternatives are improved to induce a change in behaviour. This facilitates the creation of meaningful trade-offs or tipping points but introduces endogeneity. This source of endogeneity was largely ignored until Train and Wilson (T&W) proposed a full information maximum likelihood (FIML) solution that can be implemented with simulation. In this article, we propose a limited information maximum likelihood (LIML) approach to address the SP-off-RP problem using a method which does not need simulation, can be applied with standard software and uses data that is already available for the stated problem. The proposed method is an application of the control-function (CF) method to correct for endogeneity in discrete choice models, using the RP attributes as instrumental variables. We discuss the theoretical and practical advantages and disadvantages of the CF and T&W methods and illustrate them using Monte Carlo and real data. Results show that, while the T&W method may be more efficient in theory, it may however fail to retrieve consistent estimators when it does not account properly for the data generation process if, e.g., an exogenous source of correlation among the SP choice tasks exists. On the other hand, the CF is more robust, i.e. less sensitive, to the data generation process assumptions, and is considerably easier to apply with standard software and does not require simulation, facilitating its adoption and the more extensive use of SP-off-RP data.

Keywords: Stated-preference, revealed preference, SP-off-RP, endogeneity.

1. INTRODUCTION

Stated Preferences (SP) methods are an important tool for preference elicitation because they control the variability of the data and the experimental design, may allow the presentation of non-existing alternatives and are inexpensive. To maximize the realism and the information gathered from the experiment, the choice set presented is often customized for a respondent based on their earlier revealed preferences (RP). In a seminal paper, Train & Wilson (2008) summarise the issues with endogeneity that may arise with different types of customised designs.

A very popular approach consists of using pivoting around an RP choice (Hensher, 2010), where a few hypothetical alternatives are presented that involve increases and decreases around the real-world choice for all alternatives. Train & Wilson (2008) show that pivoting can avoid the inconsistency caused by endogeneity if the model is estimated with fixed parameters, the changes are made to all alternatives irrespective of the choice observed in the RP data, and the reference alternative is not presented unchanged to the respondent labelled as "your recent choice". However, most applications of the pivoting approach do consider presenting the recent choice by itself and/or use random coefficients, cases in which a correction would be needed to attain consistency. Besides, even if the pivoting is properly built it does not guarantee that the trade-offs are tailored to create tipping points that will allow an analyst to understand at what point the decision maker will move away from the RP choice. Additionally, while the approach is suitable for unlabelled choice settings, such as route choice, it cannot easily be used for labelled settings, such as mode choice, as the attributes and values of those attributes vary across the different alternatives.

An alternative approach consists of using adaptive design methods where the original choice set is reproduced in the SP setting but where the attribute levels are changed as a function of the RP choice. The variables of the alternative chosen in the RP setting are made "worse", while those of the remaining alternatives are made "better". Examples of such approaches include the adaptive SP approaches of Shinghal and Fowkes (2002) and the adaptive conjoint of Allenby et al (1995), and more recently the SP-off-RP approach of Train and Wilson (2008; 2009).

Adaptive SP approaches have long been known to create endogeneity issues. Since the choice set presented to an individual depends on his/her RP choices, and therefore on that individual's preferences, an estimation method that neglects this endogenous setting would result in inconsistency. Correction approaches exist but they are difficult to apply. The work of Train and Wilson (2008), hereafter referred to as T&W, offered a Full Information Maximum Likelihood (FIML) solution for the specific case of SP-off-RP data that can be implemented with simulation. This seminal work has been widely cited but rarely applied, arguably because of its various practical difficulties. As a result, SP-off-RP surveys have also failed to be widely applied. Our work seeks to remedy this situation.

In this article we propose a Limited Information Maximum Likelihood (LIML) approach, based on the Control Function (CF) method, to address the SP-off-RP problem. The proposed CF method does not need simulation, can be applied with standard software and uses data that is already available for the stated problem. We discuss the theoretical and practical advantages and disadvantages of the CF and T&W methods and illustrate them using Monte Carlo and real data. Results show that, while the T&W method may be more efficient in theory, it may however fail to retrieve consistent estimators when it does not account properly for the data generation process if, e.g., an exogenous source of correlation

among the SP choice tasks exists. On the other hand, the CF is more robust, i.e. less sensitive, to the data generation process assumptions and is considerably easier to apply with standard software, facilitating its adoption and the more extensive use of SP-off-RP data.

The remainder of this paper is organized as follows. The next section presents the SP-off-RP design approach and the T&W correction for endogeneity. We follow this in Section 3 by our proposed CF method. Empirical illustrations follow in Section 4 using simulated data and Section 5 using real data. Finally, we present some conclusions and directions for future work.

2. SP-OFF-RP DATA AND TRAIN & WILSON CORRECTION FOR ENDOGENEITY

The SP-off-RP problem is a two stage process. The first stage corresponds to an RP choice that occurs in the real environment in which individual n chooses that alternative i_n^{RP} among the choice set C_n^{RP} which has the maximum utility $U_{in}^{RP} = V_{in}^{RP} + \varepsilon_{in}^{RP} = \sum_k \beta_k x_{ink}^{RP} + \varepsilon_{in}^{RP}$.

This utility depends on a systematic part V_{in}^{RP} and random part ε_{in}^{RP}. The systematic part is formed as a (typically linear) function of attributes x_{ink}^{RP} with, coefficients β_k. The random part is assumed to be independently and identically distributed Extreme Value I, with location zero and scale μ^{RP}. Under these assumptions, the probability that individual n would choose alternative i in the RP experiment will be $P_n^{RP}(i)$, as shown in Eq. (1).

$$
\begin{aligned}
&U_{in}^{RP} = V_{in}^{RP} + \varepsilon_{in}^{RP} \qquad \varepsilon_{in}^{RP} \; iid \; \mathrm{EV}\left(0, \mu^{RP}\right) \\
&y_{in}^{RP} = 1\left[U_{in}^{RP} \geq U_{jn}^{RP} \, \forall j \in C_n^{RP} \right] \\
&i_n^{RP} : \text{RP choice of individual } n \\
&P_n^{RP}(i) = \frac{e^{\mu^{RP} V^{RP}(x_{in}, \beta)}}{\displaystyle\sum_{j \in C_n^{RP}} e^{\mu^{RP} V^{RP}(x_{jn}, \beta)}}
\end{aligned}
\qquad (1)
$$

where y_{in}^{RP} takes value 1 if individual n chooses alternative i in the RP experiment.

In the second stage of the SP-off-RP choice process, the researcher builds an SP experiment which considers a choice set C_n^{SP} that includes i_n^{RP}, the alternative chosen in the RP experiment, and a set (possibly a subset) of other alternatives from C_n^{RP}.

The definition of the SP choice set C_n^{SP} has varied among the SP-off-RP applications. While Train & Wilson (2008) assumed that $C_n^{SP} = C_n^{RP} = C_n$, van Cranenburgh et al. (2014) built the SP choice set as the mix of: one or two variations of the alternative chosen in the RP stage; two alternatives randomly drawn from the non-chosen RP alternatives; and an opt-out option. We will follow the approach used by Train & Wilson (2008) to avoid other potential sources of endogeneity resulting from occasionally presenting twice the RP chosen

alternative and to avoid potential sources of hypothetical bias resulting from using a choice-set that is different from that of the RP stage.

The attributes of the resulting set of SP alternatives for each choice task t are built as variations of their values in the RP experiment in a way that tries to induce a choice change. For example, the attribute x_{jnkt}^{SP} of alternative j and individual n in the SP choice task t could be calculated as follows:

$$x_{jnkt}^{SP} = \gamma_{jnkt} x_{jnk}^{RP}, \tag{2}$$

where γ_{jnkt} is an endogenous scalar defined by the researcher such that, if $\beta_k < 0$,

$$\gamma_{jnkt} \begin{cases} > 1 \text{ if } j = i_{t_n^{RP} n} \\ < 1 \text{ if } j \neq i_{t_n^{RP} n} \end{cases}. \tag{3}$$

for given n and t. This implies that for undesirable attributes (i.e. $\beta_k < 0$), the level of the attribute is increased for that alternative which was chosen in the RP choice, while it is decreased for all other alternatives. For desirable attributes, the opposite applies. The net outcome of this is that the alternative chosen in the RP setting becomes less attractive, while those that were not chosen, become more attractive.

The SP-off-RP method described in Eqs. (1), (2) and (3) has two main goals. The first is to reduce the sources for hypothetical bias by presenting alternatives that are tailored to each respondent's reality both in terms of the choice set and the attributes presented. The second goal is to maximize the information gathered about the preferences from each answer by presenting alternatives that have attributes that vary in the direction in which they induce a change in the respondent's RP choice.

As shown by Eq. (3), for achieving the second goal, SP-off-RP requires knowledge of the sign of each coefficient, which may be seen as a limitation of the method. However, if the researcher does not know or guesses erroneously the sign of the coefficient, it would only happen that this advantage of the SP-off-RP method would be lost, leading to a loss of efficiency. Furthermore, we think it is safe to say that for a large majority of the practical examples in transportation and other fields, the sign of the coefficient can be correctly guessed. For cases in which the sign is truly unknown (and the researcher is really interested in using SP-off-RP to investigate the problem), a pilot stage of RP or classical SP data may be used to gain preliminary knowledge. Of course, this is no different from the vast use of efficient SP designs across numerous disciplines, where these not only use priors on the sign but also values of the coefficients (cf. Rose & Bliemer, 2014).

Nevertheless, the potential gains in realism and information of the SP-off-RP method comes with a caveat. Under the framework described in Eqs. (1), (2) and (3) it is plausible to think that some of the sensitivities/preferences inherent to the decision maker will transfer from the RP to the SP experiment. This effect will occur for the same reason that we assume realism increases by building SP experiments that are closely related to RP. As this transfer is not observed, it would imply that some part of the unobserved attributes of each alternative are transferred from the RP to the SP setting. This transferability can be accommodated by considering that some part of ε_{jn}^{RP} will appear also in the utility of the alternative of the SP choice model. Under that framework, the SP choice model could be written as follows

$$U_{jnt}^{SP} = V\left(x_{jnt}^{SP}, \beta\right) + \rho \varepsilon_{jn}^{RP} + \varepsilon_{jnt}^{SP} \qquad \varepsilon_{jn}^{RP}, \varepsilon_{jnt}^{SP} \ iid \ \text{EV} \mu^{RP}, \mu^{SP}$$

$$y_{jnt}^{SP} = 1\left[U_{jnt}^{SP} \geq U_{lnt}^{SP} \forall l \in C_n\right] \qquad . \qquad (4)$$

$$i_{nt}^{SP} : \text{SP Choice of individual } n \text{ on task } t$$

The coefficient ρ corresponds to the fraction of the RP error that was transferred or carried over to the SP experiment. Following van Cranenburgh et al. (2014) generalization of the SP-off-RP model, the coefficient ρ in Eq. (4) allows a situation where the individual may neglect some or even all the unobserved utility he or she had previously assigned to an alternative in the RP choice. In this way, Train and Wilson (2008) implicitly assume that $\rho = 1$, something that is not necessarily true, a point we return to below.

Note that, under this framework, the RP (Eq. 1) and the SP (Eq. 4) choices share the structural parameters of interest β_k, calling for a joint estimation. Nevertheless, the challenge from having mixed RP and SP data in this case is not only that the scale between both models may differ, as in the classic RP/SP framework analysed by Ben-Akiva et al. (1994) and Bradley and Daly (1997), but also the problem of endogeneity.

Whenever $\rho \neq 0$, endogeneity problems will arise because x_{jnt}^{SP} and hence $V\left(x_{jnt}^{SP}, \beta\right)$ will be correlated with ε_{jn}^{RP} as γ_{jnkt} depends on the RP choice and, thus, on ε_{jn}^{RP} (cf. Eqs. 1-4). Therefore, consistency in the estimation of the model coefficients will not be achieved if an analyst ignores this transfer of error, i.e. estimates a model either on just the SP data or jointly on the SP and RP data, without creating a proper link between the two. This thus potentially leads to inconsistency in the model parameters.

The choice model depicted in Eq. (4) also considers an idiosyncratic[1] error term ε_{jnt}^{SP}, which corresponds to exogenous unobserved factors that affect the utility in the SP experiment and changes over time.

The seminal work of Train and Wilson (2008) put forward a solution to the endogeneity problem arising in the model in Eq. (4). First, considering that the idiosyncratic error ε_{jnt}^{SP} is distributed iid Extreme Value I $\left(0, \mu^{SP}\right)$, and conditioning on ε_{jn}^{RP}, the choice model in Eq. (4) can written as a logit. Then, using that logit model as the kernel, the probability of alternative j in the tth SP scenario, conditional on i being chosen in the RP setting, is given by a mixed logit model as:

$$P_{nt}^{SP}\left(j \mid i_n^{RP}\right) = \int \frac{e^{\mu^{SP}\beta' X_{jnt}^{SP} + \rho \mu^{SP} \varepsilon_{jn}^{RP}}}{\sum_{l \in C_n} e^{\mu^{SP}\beta' X_{lnt}^{SP} + \rho \mu^{SP} \varepsilon_{ln}^{RP}}} f\left(\varepsilon_n^{RP} \mid \beta' X_{i_n^{RP}}^{RP} + \varepsilon_{i_n^{RP}}^{RP} > \beta' X_{ln}^{RP} + \varepsilon_{ln}^{RP} \quad \forall l \neq i_n^{RP}\right) d\varepsilon_n^{RP} \quad ,$$

(5)

where the integration is over the density of ε_n^{RP}, conditional on i_n^{RP} being chosen in the RP setting.

Joint estimation on the RP and T SP choices made by each induvial n then maximizes:

---

[1] Train and Wilson (2008) called this error term "quixotic"

$$P_n(Y_n) = \int \prod_{t=1}^{T} \frac{e^{\mu^{SP}\beta'X_{jnt}^{SP}+\rho\mu^{SP}\varepsilon_{nj}^{RP}}}{\sum_{l\in C_n} e^{\mu^{SP}\beta'X_{lnt}^{SP}+\rho\mu^{SP}\varepsilon_{ln}^{RP}}} f\left(\varepsilon_n^{RP} \mid \beta'X_{i_n^{RP}}^{RP}+\varepsilon_{i_n^{RP}}^{RP} > \beta'X_{ln}^{RP}+\varepsilon_{ln}^{RP} \quad \forall l \neq i_n^{RP}\right) d\varepsilon_n^{RP} \frac{e^{\beta'X_{i_n^{RP}}^{RP}+\varepsilon_{i_n^{RP}}^{RP}}}{\sum_{l\in C_n} e^{\beta'X_{ln}^{RP}+\varepsilon_{ln}^{RP}}} \quad (6)$$

where Y_n corresponds to the vector containing the RP choice and the sequence of choices in the T SP tasks. Train and Wilson (2008) consider the likelihood shown in Eq. (6) but with $\rho=1$. As it was stated before, following van Cranenburg et al (2014), we consider a further level of flexibility by not fixing ρ to 1, allowing for different levels of error transfer. The estimate for μ^{SP} is then a scale parameter for SP data while ρ gives an indication of how much of the RP error is transferred to the SP setting.

The actual way to produce draws from ε_n^{RP} is not computationally difficult albeit tedious to implement given that they come from an extreme value distribution with a shifted mean (by the negative logarithm of the probability of the RP choice) for the alternative chosen in the RP setting and a truncated extreme value distribution for those alternatives not chosen in the RP setting (truncated by the differences in the deterministic utilities for i and l plus the error term for alternative i). Full details on this are given in Train and Wilson (2008). Even if the model the analyst wishes to estimate is Multinomial Logit (MNL), simulation-based estimation is required with this method.

3. PROPOSED CONTROL FUNCTION CORRECTION FOR THE SP-OFF-RP PROBLEM

In this article, we propose a method to correct for the endogeneity problem in the SP-off-RP problem taking advantage of the fact that RP attributes are suitable instrumental variables for the SP attributes, thus allowing us to apply the control-function (CF) approach. The method we propose has the advantage that, as all LIML methods, it is more robust (i.e. less sensitive) to misspecifications of the underlying distributional assumptions, but with a potential cost in efficiency (Wooldridge, 2010, section 9.3.1). Besides, the proposed method is much easier to apply, facilitating its adoption and more extensive use.

Robustness in this case means that the proposed CF method is less sensitive to misspecification in the model's errors. This means that various error structures would result in the proposed method providing consistent estimators of the model parameters, while the state-of-the-art T&W method would only achieve consistency for the specific joint error distribution assumed in Eqs. (1) and (4). The value of this robustness is reflected in the Monte Carlo experiments presented in Section 4 and the real data case study analysed in Section 5, which suggests that T&W method provides inconsistent estimators when it neglects a source of common exogenous variation among SP choice tasks, while the proposed CF method is immune to that problem.

The CF method was originally proposed by Heckman (1978), with later contributions by Rivers and Vuong (1998), Petrin and Train (2002), Guevara and Ben-Akiva (2006, 2012), Guevara (2005, 2010, 2015, 2018), Terza et al. (2008) and Terza (2018), who termed it the "two-stage residual inclusion estimation". The application of the CF for the correction of endogeneity requires an instrumental variable for each endogenous variable of the model. This instrumental variable must be, at the same time, sufficiently correlated with the respective endogenous variable and independent of the error term of the model.

Because of Eq. (2), the endogenous variables in this case are each x^{SP}_{jnkt} built as a variation of the respective x^{RP}_{jnk}, because γ_{jnkt} depends on the RP choice and hence on the transferred ε^{RP}_{jn}. Also because of Eq. (2), x^{RP}_{jnk} makes a proper instrument for x^{SP}_{jnkt} because they are correlated by construction, and under the assumption that the RP model was correctly specified, x^{RP}_{jnk} is independent of ε^{RP}_{jn}, which is the source of endogeneity in the SP model. Therefore, we can consistently estimate the SP model alone if we correct for the endogeneity using the two stages control-function (2SCF) approach[2].

To explain the proposed 2SCF method, consider a binary mode choice model between car and bus, and with only two x^{RP}_{jnk} attributes: travel time and cost. Since people would like to travel for a shorter time and paying a smaller amount of money, $\beta_t < 0$ and $\beta_c < 0$ respectively. Then, if, for example, individual n chooses the bus in the RP experiment, the attributes of the SP experiment will be built to induce a behaviour change, worsening the bus and improving the car. This could be achieved by, for example, creating the following SP attributes shown in Eq. (7).

$$
\begin{aligned}
Time^{SP}_{Bus\_n} &= 1.1 * Time^{RP}_{Bus\_n} \\
Cos\,t^{SP}_{Bus\_n} &= 1.3 * Cos\,t^{RP}_{Bus\_n} \\
Time^{SP}_{Car\_n} &= 0.8 * Time^{RP}_{Car\_n} \\
Cos\,t^{SP}_{Car\_n} &= 0.7 * Cos\,t^{RP}_{Car\_n}
\end{aligned}
\tag{7}
$$

In this case the scalars γ_{jnk} could have been fully predefined (e.g. 1.1 for time for the RP chosen alternative in this case) or randomly drawn from a predefined set of scalars that are larger or smaller than one, as appropriate. We will follow the later approach in the Monte Carlo experiments developed later, on Section 4.

Then, the 2SCF correction is applied in two stages. In the first stage, each endogenous x^{SP}_{jnkt} should be regressed on their respective instrument and the controls, to retrieve a residual that will capture the part of x^{SP}_{jnkt} that was correlated with the ε^{RP}_{jn} error of the RP model that was transferred to the SP task. Because of Eq. (7), the instrument for each x^{SP}_{jnkt} is its respective x^{RP}_{jnk}.

Because in this example we have more than one endogenous variable, the regression must be made not only on the respective instrument, but also on the instruments of all the other endogenous variables, which act as controls. For the stated example, this means estimating the following two regressions shown in Eq. (8) by ordinary least squares (OLS), stacking the information from all available alternatives.

---

[2] Please refer to Wooldridge (2010) for a formal demonstration of the consistency of the 2SCF method.

$$Time_{jnt}^{SP} = \alpha_T + \alpha_{TT}Time_{jn}^{RP} + \alpha_{TC}Cost_{jn}^{RP} + \delta_{jnt}^{Time}$$
$$Cost_{jnt}^{SP} = \alpha_C + \alpha_{CT}Time_{jn}^{RP} + \alpha_{CC}Cost_{jn}^{RP} + \delta_{jnt}^{Cost}.$$

(8)

Then, in the second stage, the residuals of these regressions $\hat{\delta}_{jnt}^{Time}, \hat{\delta}_{jnt}^{Cost}$ are added to the SP choice model to control for endogeneity, such that the systematic part of the utility takes the form shown in Eq. (9).

$$V_{jnt}^{SP} = ASC_j + \beta_{time}Time_{jnt}^{SP} + \beta_{cost}cost_{jnt}^{SP} + \theta_T\hat{\delta}_{jnt}^{Time} + \theta_C\hat{\delta}_{jnt}^{Cost}$$

(9)

Under the setting described, the proposed method will allow the consistent estimation of the model coefficients $ASC_j, \beta_{time}, \beta_{cost}$ shown in Eq. (9), up to a scale (Guevara and Ben-Akiva, 2012).

For this method to be valid, we would formally need to assume that ε_{jn}^{RP} is normally distributed (see, e.g. Guevara, 2015), implying that the RP model is a Probit instead of a Logit. However, if the goal is to achieve consistency, the Probit model could be safely approximated by the Logit model shown in Eq. (1), provided the *iid* property is maintained (Lee, 1982; Ruud, 1983).

The standard errors of the estimators obtained with the 2SCF method cannot be obtained directly from the inverse of the Fisher Information Matrix because the residual used in the second stage is an estimated regressor. This problem can be solved using the Delta method (Karaca-Mandic and Train, 2003) or a nonparametric approach, such as the bootstrap (Petrin and Train, 2002). Nevertheless, a statistical test of the null hypothesis that $\hat{\theta}_T$ and $\hat{\theta}_C$ are equal to zero, obtained directly from the information matrix, can be used as a test for the absence of endogeneity because the estimated regressor problem disappears under the null (Rivers and Vuong, 1988).

The need for estimating the standard errors indirectly can be circumvented using what Train (2009, Ch. 13.5) terms the Maximum Likelihood (ML) approach of the CF method (MLCF hereon), which consists of considering the stages described in Eqs. (8) and (9) simultaneously. Such an approach may also increase the efficiency of the estimators, compared to those of the 2SCF version, but at the cost of reducing robustness or generality. The loss of robustness is related with that, while 2SCF requires a specification of the conditional distribution of $\varepsilon_{jn} = \rho\varepsilon_{jn}^{RP} + \varepsilon_{jnt}^{SP}$ given δ_{jnt}, MLCF requires specifying their joint distribution. The problem is that, while there are many joint distributions that have a specific conditional distribution, each joint distribution implies only a certain conditional distribution (see e.g. Train, 2009, Ch. 13.5 for discussion specific for CF and Wooldridge, 2010, section 9.3.1, for a general one).

It should be noted however that, even though MLCF is often more efficient than the 2SCF method, it is still a LIML approach and as such is less efficient than the FIML method proposed by T&W for the SP-off-RP problem under study. At the same time however, both CF approaches are more robust (i.e more general or less sensitive) than the FIML one in the sense that they require fewer distributional assumptions to attain consistency.

In the Monte Carlo experiments reported in Section 4 we will consider the two stages version of the CF method (2SCF) since the standard errors for the statistics of interests will be calculated from the repetitions of the experiments, and efficiency is a controlled issue in that case. However, in the application with real data reported in Section 5, we will use the MLCF version, not only to facilitate the calculation of the standard errors directly from the information matrix, but also to make the best of a database that has very few observations and low levels of variance.

Compared to the T&W method, the proposed 2SCF will be further less efficient, not only because it is a LIML estimator, if the estimation of the SPs alone in the 2SCF misses the information on the RP choice, which are included in T&W's approach. This limitation can be easily handled with minimum increase of the model complexity by including the RP observations, but accounting for the change of scale, as shown by Ben-Akiva et al. (1994) and Bradley and Daly (1997). This approach will be used later, both in the Monte Carlo and the real data experiments studied in sections 4 and 5. It should be noted, however, that it is not necessarily always convenient to include the RP data in the estimation. One case in which it would not be appropriate to do it so would be when the RP attributes are self-reported, because that may induce a different type of endogeneity problem (see, e.g Van Cranenburgh et al. 2004).

Another reason why the basic implementation of the 2SCF method may be less efficient than the T&W method is because that the lack of random components (and hence integration) may neglect the potential correlation across SP tasks. The T&W approach relies on capturing correlation between each SP-off-RP scenario and the RP scenario on which they were built. SP surveys typically present respondents with multiple scenarios (generally between 8 and 16) and there is ample evidence of correlation between the responses to individual scenarios coming from the same respondent, where this is of great benefit in retrieving random heterogeneity across respondents. However, there is clearly also the possibility that there is correlation in the error terms between the SP scenarios which is specific to the SP data and not linked to the original RP data. A main reason for such correlation is that the behaviour in hypothetical settings may differ from that in real world settings. If such SP-specific correlation exists, then the FIML approach used by T&W runs the risk of misattributing this correlation between SP scenarios to correlation between the RP and SP scenarios, a point we illustrate in our empirical work. Furthermore, the limitation of the 2SCF in the sense of neglecting any existing correlation among SP choices may only compromise efficiency, but not consistency of the estimators and if efficiency is an issue, it can be handled in that case by extending the 2SCF approach to account for correlation across the different SP choices using a mixed distribution (Train, 2009, Ch. 6), at the cost however of increasing the method's complexity. We will explore the impact of this improvement in our empirical work.

The proposed CF method is easier to apply and that, we think, may facilitate its adoption and allow a more extensive use of SP-off-RP approaches. This does not mean that the CF method can be used with standard software immediately but, as described before, by the combination of proper applications of standard software in ways that are different for each database. The main reason the CF approach is "easier" comes from the fact that the method does not require ad-hoc simulation procedures and is therefore less computationally

demanding, and also less prone to normalization pitfalls and to empirical false identification problems.

Summarizing, while the proposed 2SCF method may be less efficient than T&W method to address the endogeneity problem on SP-off-RP data, it has two important advantages. The first is that, as all LIML approaches, the 2SCF is more robust to the underlying distributional assumptions of the model. This advantage manifests itself in this case, as we will show later, in that the T&W approach can misattribute any correlation between the SP scenarios to correlation between RP and SP scenarios, while the CF is not affected by that problem. The second relative advantage of the proposed CF approach is that it is much easier to implement than the T&W approach and can be applied using standard software if it is estimated in two stages.

4. MONTE CARLO EXPERIMENT CONTRASTING THE CF AND T&W APPROACHES UNDER VARIOUS DATA GENERATION PROCESSES

In this section, we report a Monte Carlo Experiment to contrast the performance of the proposed Control Function (CF) method with the current state of the art for addressing endogeneity in the SP-off-RP problem under various data generation processes. The experimental setting is designed to illustrate the main advantages and limitations of both approaches.

The data generation process for the Revealed Preference (RP) experiment is a trinomial mode choice model, where 250 individuals must choose solely on the basis of travel time ($Time_{jn}^{RP}$) and travel cost ($Cost_{jn}^{RP}$), with alternative specific constants set to zero and using the population coefficients shown in Eq. (10).

$$U_{jn}^{RP} = -1.0 * Time_{jn}^{RP} - 0.5 * Cost_{jn}^{RP} + \varepsilon_{jn}^{RP}. \tag{10}$$

$Time_{jn}^{RP}$ and $Cost_{jn}^{RP}$ values were drawn from a random uniform between 1 and 3. The error ε_{jn}^{RP} was drawn from a standard Normal distribution to facilitate the variance control among experiments, as explained later. Although the error term in data generation was distributed normal, logit family models were used in estimation, in line with Lee's (1982) and Ruud's (1983) results.

Eight[3] SP choice tasks t were built for every individual n in the sample by shifting the attributes of the RP experiment following the SP-off-RP idea. The multipliers used to shift travel time and cost were randomly drawn for each individual and alternative from a uniform distribution between 1.1 and 1.4 for the alternative that was chosen in the RP, and between 0.6 and 0.9 for the remaining two alternatives.

The utility of the model used to build each SP choice task t was then given by

---

[3] As with every SP experiment, the number of SP choice tasks in a SP-off-RP setting plays a role in the efficiency attainable by the model because more tasks imply more observations per individual. The Monte Carlo experiments considered eight choice just to resemble what may be used in a typical application.

$$U_{jnt}^{SP} = -1Time_{jnt}^{SP} - 0.5Cost_{jnt}^{SP} + \gamma_1 e_{jn}^{SP} + \gamma_2 e_{jnt}^{SP} + \gamma_3 \varepsilon_{jn}^{RP}, \tag{11}$$

where e_{jnt}^{SP} is an exogenous standard Normal error term that is only present in the SP choices and varies across tasks t; e_{jn}^{SP} is another exogenous standard Normal error term repeated across SP choices for each individual n; and ε_{jn}^{RP} is the RP error term.

This data generation process allows for two types of SP specific exogenous errors. e_{jnt}^{SP} is an idiosyncratic term that affects the SP utility and changes over time in the same way as in T&W's model, as shown in Eq. (4). The second SP specific exogenous error e_{jn}^{SP} allowed captures SP specific factors that do not vary by individual and are therefore shared between their choice tasks, causing some type of correlation among them that is not related to the RP choice. Borrowing from the panel data literature, we will denominate the error term e_{jn}^{SP} as a random effect.

The Monte Carlo experiments developed in this section, and the real data application in Section 5, will show that neglecting the random effect in T&W's FIML approach may preclude the consistent estimation of the model parameters, while the LIML approach to the problem proposed in this paper may be robust to this type of distributional misspecification.

Four experimental cases for the data generation process were defined, based on the coefficients $\gamma_1, \gamma_2, \gamma_3$ shown in Eq. (11).

Case 1 $\gamma_1 = 0, \gamma_2 = 1, \gamma_3 = 1$: This case corresponds to the data generation process assumed by Train and Wilson (2008), which leads to the likelihood shown in Eq. (6) with $\rho=1$. The RP error ε_{jn}^{RP} is fully $(\gamma_3 = 1)$ transferred to the SP choices, causing endogeneity, and there is an idiosyncratic exogenous error e_{jnt}^{SP} that changes between choice tasks t, producing also a change in scale between RP and SP models.

Case 2 $\gamma_1 = 0, \gamma_2 = 1/\sqrt{2}, \gamma_3 = 1/\sqrt{2}$: This case corresponds to the data generation process for the model considered by Van Cranenburgh et al (2014), where not necessarily all the RP error ε_{jn}^{RP} is transferred to the SP choice tasks, but still causing endogeneity. The value considered for $\gamma_2 = \gamma_3 = 1/\sqrt{2}$ is used to guarantee that the variance in each SP experiment is the same as that of the RP experiment, to avoid differences that may be attributable to model precision. Note that this also implies that no scale differences should exist between the RP and the SP experiments, provided they are properly specified.

Case 3 $\gamma_1 = 1/\sqrt{3}, \gamma_2 = 1/\sqrt{3}, \gamma_3 = 1/\sqrt{3}$: This case, the most general studied in this paper, adds to Case 2 the possibility of having random effect SP errors e_{jn}^{SP} that are shared across SP choice tasks. The value of $1/\sqrt{3}$ is again used to guarantee that the variance in each SP choice task is equal to 1. Since ε_{jn}^{RP} is partially transferred to the SP choices, this model also suffers from endogeneity.

Case 4 $\gamma_1 = 1/\sqrt{2}, \gamma_2 = 1/\sqrt{2}, \gamma_3 = 0$: In this case, it is assumed that there is no endogeneity because the RP error is not transferred to the SP choice tasks. However, there is a random effect exogenous error e_{jn}^{SP} that is shared by the SP choice tasks together with an idiosyncratic task specific error e_{jnt}^{SP}. As we will show, this setting can result in confounding sources of correlation among SP choice tasks that may result in an undesired source of endogeneity when the model is estimated with a wrongly specified FIML model.

For each of the 4 Cases considered for the data generation process, the following 11 models were estimated:

1. **RP**: RP model estimated considering all the explanatory variables described in Eq. (10). This model is used as a benchmark of a model that should not suffer from endogeneity.
2. **RP/SP**: Joint estimation of the choice model considering both the RP and the SP-off-RP data, neglecting however the fact that the attributes may be endogenous. This model is used as the do-nothing benchmark.
3. **RP/SPσ**: RP/SP model where an error component mixture is used to account for the correlation among SP choice tasks due to a potential random effect[4]. This model neglects the potential existence of endogeneity problem but allows for the correlation among the SP choice tasks.
4. **RP/SPμ**: RP/SP model where the potential scale difference between the RP and SP choice experiments is accounted for, following the approach proposed by Ben-Akiva et al. (1994) and Bradley and Daly (1997). The potential endogeneity problem is however neglected as is the correlation between SP tasks.
5. **RP/SPμσ**: RP/SP model where the scale and the error component to account for a random effect are considered. The potential endogeneity problem is neglected.
6. **RP/SP_CF**: Joint estimation of the choice model considering both the RP and the SP-off-RP data and correcting for the potential endogeneity problem using the proposed CF method, as discussed in Section 3.
7. **RP/SP_CFμ**: Joint estimation of the RP and SP-off-RP model correcting for endogeneity using the CF method, but now also including the scale correction.

---

[4] In this and in later models using σ, this refers to an additional N(0,σ) simulated error term which is distributed identically but independently across individuals and across alternatives, inducing correlation across SP observations for the same individual without introducing either correlation or heteroscedasticity across alternatives.

8. **RP/SP_CFμσ**: Joint estimation of RP and SP-off-RP correcting for endogeneity using the CF method, including the scale correction and an error component to account for the correlation among SP choice tasks due to a random effect.
9. **T&W**: The RP and SP-off-RP models are estimated together using the Train and Wilson (2008) correction method to account for SP-off-RP endogeneity, as described shown in Eq. (6), assuming a full error transfer with ρ=1.
10. **T&Wρ**: T&W model but, following van Cranenburgh et al (2014), allowing for a partial transfer of the RP error into the SP choice tasks, as shown in Eq. (6), by estimating ρ.
11. **T&Wρσ**: T&W model considering partial transfer of RP error and an error component mixture that is used to account separately for the correlation among SP choice tasks due to a random effect.

These 11 models are analysed in terms of their empirical finite sample properties using the Monte Carlo outcomes. Because the correction of endogeneity in discrete-choice models produces a change in the scale of the model (Guevara and Ben-Akiva, 2012), we study the sampling distribution of the ratio $\hat{\beta}_t / \hat{\beta}_c$ instead of the $\hat{\beta}_t$ and $\hat{\beta}_c$ themselves. The analysis was developed by repeating each experiment 100 times and analysing the empirical distribution of the estimates obtained from those repetitions to calculate the following four statistics:

- **% Bias:** Mean percent empirical finite sample bias of $\hat{\beta}_t / \hat{\beta}_c$ across the 100 repetitions of the Monte Carlo experiments, which can be calculated in this Monte Carlo experiment because the population parameters are known. An endogenous model should show a large % Bias and a model that successfully corrects for endogeneity should have a small value for % Bias.
- **p-value:** p-value of the asymptotic t-test for the hypothesis that the estimated ratio is equal to the population value $\beta_t / \beta_c = 2$. p-values above, e.g., 0.05 are interpreted as a successful correction of the potential endogeneity problem.
- **Time (sec)**: Average estimation time in seconds.
- **LL:** Average log-likelihood across the 100 repetitions. Less negative values indicate better fit, but sample size related and does not account for number of parameters. Sample size is 250 for the RP and 2250 for all other cases that combine RP and SP choices.
- **adj. ρ^2:** goodness of fit measure adjusted for the number of parameters (see e.g. Ben-Akiva and Lerman, 1985). Higher values indicate better fit but still sample size related.

For illustrative purposes, the results from Table 1 are also shown in Figure 1. The vertical axis in Figure 1 corresponds to the estimated ratio $\hat{\beta}_t / \hat{\beta}_c$ and the methods are ordered along the horizontal axis. The dashed horizontal line marks the true population value of the ratio $\beta_t / \beta_c = 2$ and the boxplots are built from the estimates obtained from the 100 repetitions of the Monte Carlo experiment. The bold horizontal line on each boxplot corresponds to the median of the empirical sampling distribution and the grey diamond

marks the respective sample mean. The upper and lower box shows the 25% and 75% percentile, and the whiskers are limited by 1.5 times the inter-quartile range.

Table 1: Results for Monte Carlo Experiments

Model	Case 1					Case 2					Case 3					Case 4				
	% Bias	p-value	time (sec)	LL	adj. ρ^2	% Bias	p-value	time (sec)	LL	adj. ρ^2	% Bias	p-value	time (sec)	LL	adj. ρ^2	% Bias	p-value	time (sec)	LL	adj. ρ^2
1. RP	2	.23	<1	-195	0.28	2	.23	<1	-195	0.28	2	.23	<1	-195	0.28	2	.23	<1	-195	0.28
2. RP/SP	79	<.01	<1	-2324	0.059	33	<.01	<1	-2126	0.14	23	<.01	<1	-2090	0.15	1	.53	<1	-1935	0.22
3. RP/SPσ	40	<.01	49	-2127	0.14	23	<.01	47	-1961	0.20	12	<.01	61	-1706	0.31	-1	.51	64	-1711	0.31
4. RP/SPμ	51	<.01	1	-2304	0.067	31	<.01	1	-2122	0.14	22	<.01	1	-2087	0.15	1	.51	1	-1934	0.22
5. RP/SPμσ	26	<.01	96	-2119	0.14	22	<.01	65	-1961	0.20	13	<.01	97	-1705	0.31	<1	.78	75	-1705	0.31
6. RP/SP_CF	1	.72	1	-2239	0.094	1	.30	1	-2066	0.16	1	.25	1	-2050	0.17	2	.14	1	-1933	0.22
7. RP/SP_CFμ	1	.44	1	-2230	0.10	1	.26	1	-2065	0.16	2	.23	1	-2049	0.17	2	.14	1	-1932	0.22
8. RP/SP_CFμσ	2	.25	115	-2092	0.15	2	.14	92	-1940	0.21	1	.23	118	-1694	0.31	1	.48	125	-1704	0.31
9. T&W	2	.25	199	-2059	0.17	-1	.43	198	-1927	0.22	-10	<.01	202	-1708	0.31	-17	<.01	210	-1785	0.28
10. T&Wρ	2	.19	200	-2059	0.16	1	.16	273	-1913	0.22	-9	<.01	276	-1701	0.31	-4	.02	321	-1774	0.28
11. T&Wρσ	3	.03	296	-2058	0.16	3	.01	392	-1912	0.22	2	.15	379	-1682	0.32	1	.25	459	-1705	0.31

Consider first the RP model. The results are the same for the four cases analysed as the RP data remains the same. Endogeneity is not present in this case and therefore the population parameters were properly recovered, showing percent empirical finite sample bias of 2% and p-values of 0.23. Estimation time was below half a second. Note also that, compared to the other models, the log-likelihood (LL) is 10 times smaller, given the smaller sample size. For the same reason, the boxplot of this model tends to be thicker than other models that successfully address the endogeneity problem using the SP-off-RP data.

Model 2, the RP/SP model without correction for endogeneity instead shows a significant percent bias and close to zero p-values for Cases 1, 2 and 3, where the endogeneity problem due to the transfer of the RP error manifests itself in different forms. Instead, for Case 4, where endogeneity is not present, the RP/SP model does recover the population parameters correctly. The same conclusions are gathered from the boxplots depicted in Figure 1. Estimation times for the RP/SP model are below one second, considerably less than T&W's estimations, because it does not involve simulation.
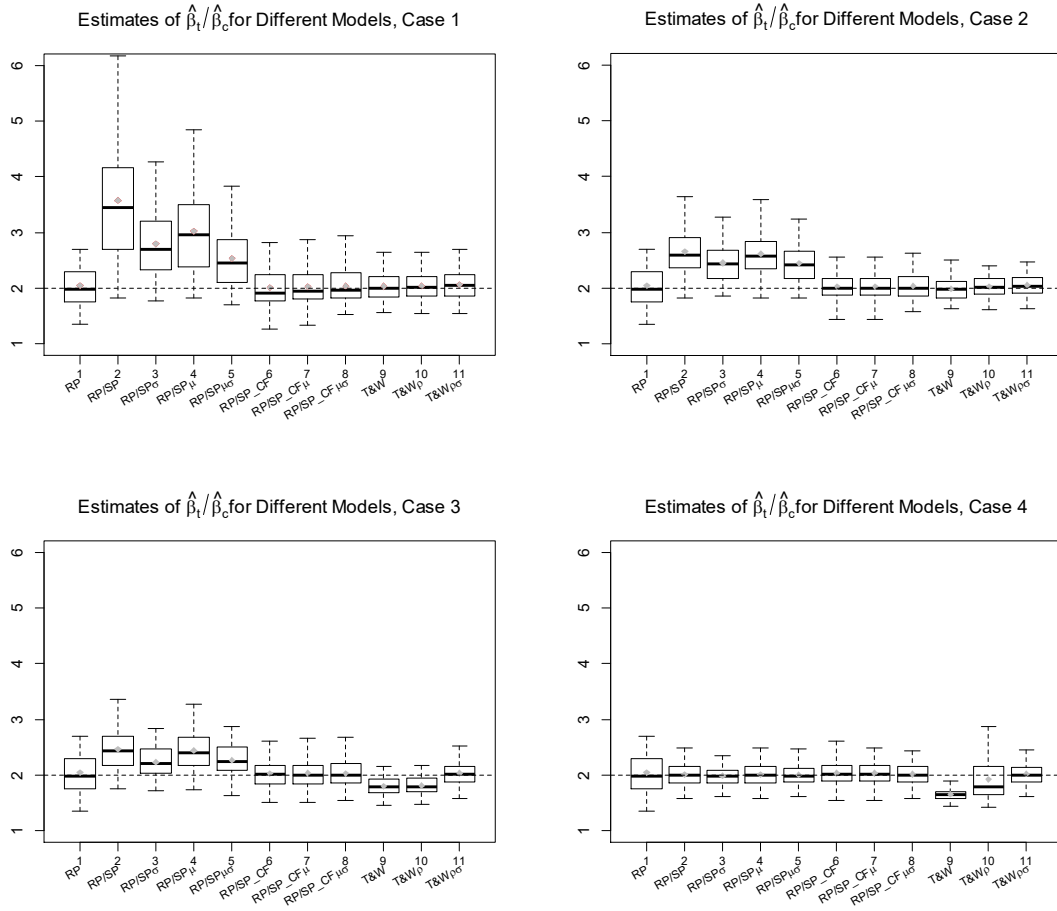
Figure 1: Boxplot of Methods to Correct for Endogeneity in SP-off-RP Data

Consider now Model 3, RP/SPσ, where an error component is added to the RP/SP model to account for the correlation between the SP choice tasks due to a potential random effect. For Cases 1, 2 and 3, the endogeneity problem remains with large percent bias and small p-values. However, the model fit does improve significantly (e.g adj. ρ^2 is almost doubled for all cases) since this model is capturing a correlation between the SP choice tasks that really exists, introduced through γ_1 and/or γ_3 in Eq. (11). For Case 4, where endogeneity was not present, this model recovers the population parameters while also achieving a better fit than the RP/SP (considering the mean adj. ρ^2 by observation). The same conclusions are gathered from the boxplots depicted in Figure 1. Finally, it can be noted that, due to the simulation needed to account for the error component, estimation time increases significantly, going now up to almost one minute.

For model 4, RP/SPμ, the results are similar. Endogeneity remains for Cases 1, 2 and 3, but is not present in Case 4. Estimation time is again close to one second since no simulation is needed and, compared to RP/SP, the fit improves for Case 1, were a change of scale does exist, but this advantage disappears for Case 4, where there is no endogeneity and the RP and SP scales are the same because in both the RP and the SP models, the error term

is standard normal. The small reduction of bias and increase in fit over the RP/SP model for cases 2 and 3 could be explained by the confounding effect of the endogeneity that is present in those cases. The same conclusions are gathered from the boxplots depicted in Figure 1.

Results for model 5, RP/SPμσ, are qualitatively the same of those of model RP/SPσ, showing a large percent bias and small p-value for cases 1, 2 and 3, where endogeneity exists, and small percent bias and large p-values for case 4, where endogeneity is not present. Model fit is significantly better than that of model RP/SP as the model accounts for a true correlation between the SP choice tasks. The same conclusions are gathered from the boxplots depicted in Figure 1.

Consider now the models estimated using the three variations of the CF method proposed in this paper to correct for endogeneity in SP-off-RP data: RP/SP_CF, RP/SP_CFμ and RP/SP_CFμσ. It can be noted that, for all cases, the population parameters were properly recovered, showing small percent bias and large p-value. The same conclusions are gathered from the boxplots depicted in Figure 1. The proposed method succeeded in correcting the endogeneity that was present in cases 1, 2 and 3, and worked well also for Case 4, where endogeneity was not present. For RP/SP_CFμ the fit improved only for Case 1, the only situation where a change of scale was indeed present. Besides, RP/SP_CFμσ did improve the fit in all cases because of the existence of correlation among SP choice tasks due to the random effect e_{jn}^{SP} and/or to the transferred RP error e_{jn}^{RP}, and the estimation time was significantly larger due to the need of simulation.

Consider finally the models estimated using the three variations of T&W method to correct for endogeneity in SP-off-RP data: T&W, T&Wρ and T&Wρσ.

The first notable result is that T&W solves the endogeneity problem for cases 1 and 2 but fails to correct for endogeneity in Case 3 and, more troubling, produces inconsistent estimators in Case 4, a situation where endogeneity was not present. Something similar occurs for T&Wρ, which fails for cases 3 and 4 with the small difference that this model implied an improvement in fit over T&W for cases 2, 3, and 4, where the RP error was indeed not fully transferred. The reason behind this result is that T&W and T&Wρ are FIML approaches that neglect the fact that, in cases 3 and 4, there is an exogenous source of correlation among the SP choices $\left(e_{jn}^{SP}\right)$, and that correlation across the SP scenarios (not related to the RP scenario) is confounded with the transferred RP error that is shared across SP choices. This makes the model fail to correct for endogeneity in Case 3 and to imply the presence of non-existent endogeneity in Case 4. For cases 1 and 2, when T&W and T&Wρ work well, the fit of these FIMIL models is better than that of the LIML model RP/SP_CF, but the difference in fit almost disappears when the correlation among SP choices and scale difference is accounted for in the RP/SP_CFμσ model. The CF correction methods of course has the more important advantage of being more robust since it still succeeds in the correction of endogeneity for cases 3 and 4, with the addition of being easier to apply and significantly faster to estimate.

The fact that the failure of T&W and T&Wρ for cases 3 and 4 is due to the confounding effect between the RP error and the random effect error that is shared by the SP choices is confirmed by the success of the model T&Wρσ. In this case, thanks to the exogenous error component being considered, both effects can be properly disentangled to

obtain consistent estimates of the model parameters, with large p-values and small percent errors for all four cases. The same conclusions are gathered from the boxplots depicted in Figure 1, where it can be noted that for cases 3 and 4, only T&Wρσ among the T&W corrections has a boxplot clearly centred on the population value.

Summarizing, the Monte Carlo experiments show that the proposed CF method to correct for endogeneity in SP-off-RP data is robust to the four cases analysed. On the other hand, the T&W approach without additional SP specific error terms fails to correct the endogeneity problem when it is confounded with exogenous sources of correlation among the SP choices, a problem that may result in inconsistent estimators even for cases in which endogeneity was not present. On the other hand, when the T&W approach works, it is more efficient than the proposed CF method. These results illustrate a trade-off between the FIML and LIML estimators, with the former being more efficient but the latter more robust to the data generation assumptions.

5. APPLICATION WITH REAL DATA

For the analysis with real data, we revisit the freight choice experiment presented by Train and Wilson[5] (2008). This dataset contains a very small sample, potentially contributing to numerous issues with low parameter significance, but we reuse it here to enable us to compare our results with the original T&W work. Our analysis on real data has three specific aims, namely:
   a) to illustrate the application of the proposed CF method to the same problem as used previously with the FIML approach of T&W;
   b) to compare performance in terms of computational cost, efficiency and consistency; and
   c) to establish the potential impact on both approaches of SP-specific correlations across choices in a real dataset.

The dataset looks at agricultural shippers in the Pacific Norwest, from the Whitman Country to Portland, using SP-off-RP. The dataset looks at 103 shippers making the RP choice of route between the following 6 options, each one described by its rate (in dollars per ton), time (in days) and reliability (as % of arrival on time):

1. truck to Pasco and barge to Portland;
2. truck to another barge port and barge to Portland;
3. rail to Portland;
4. truck to a rail terminal and rail to Portland;
5. barge to Portland;
6. other.

Each shipper was first asked to declare the alternatives that they had available and the rate, transit time and reliability of each of them.
Afterwards, the respondents faced separate SP experiments, each containing the same set of alternatives as their RP choice. The first faced them with a scenario where the rate of

---

[5] We would like to thank Kenneth Train for making the data available for use in this article.

the alternative they chose in the RP setting was shifted by X percent higher, where X was randomly selected from 10, 20, 30, 40, 50 and 60. The second and third SP tasks used a similar scheme for increases of transit time and decreases in reliability, respectively, for the RP chosen alternative. Each time changes were made only for the alternative that was chosen in the RP setting.

Table 2 summarizes the MNL estimation results of the RP experiment, followed by separate results for each SP experiment, and the results for the pooled sample of the RP and the SP experiments accounting for the traditional RP/SP scale change, but not for endogeneity. In line with T&W, we report the mean log-likelihood (LL) per observation, but in addition the adj. ρ^2 to account for the number of parameters and aid comparison across models. In all models, we use alternative 2 as the base alternative, with its ASC normalized to zero.

Table 2. No Endogeneity Correction Portland SP-off-RP Freight Data

	RP		SP1 (Change rate)		SP2 (Change time)		SP3 (Change relia.)		Pooled RP/SP	
	est.	rob.t[6]	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t
β_{Rate}	-0.125	-1.37	-0.0741	-1.51	-0.125	-1.92	-0.132	-1.93	-0.175	-1.56
β_{Time}	-0.034	-1.09	-0.0160	-0.44	0.0048	0.155	-0.0151	-0.50	-0.0417	-1.28
β_{Relia}	0.0322	2.34	0.0140	1.32	0.0143	1.28	0.00595	0.615	0.028	2.55
ASC_1	-1.742	-2.63	-0.371	-0.935	-0.271	-0.675	-0.355	-0.885	-1.113	-1.62
ASC_2	0	-	0	-	0	-	0	-	0	-
ASC_3	1.08	2.30	-1.10	-1.63	-0.882	-1.40	-0.884	-1.38	0.103	0.165
ASC_4	-0.675	-2.03	-0.645	-1.70	-0.811	-2.19	-0.838	-2.47	-1.018	-2.78
ASC_5	-0.456	-0.435	-0.637	-0.705	-0.0904	-0.095	-0.0518	-0.04	-0.608	-0.455
ASC_6	-0.596	-0.425	-0.539	-0.485	-0.245	-0.205	-0.0850	-0.07	-0.915	-0.465
μ_{RP}									1.00	-
μ_{SP}									0.428	1.79
Value of Time	0.273	0.875	0.216	0.455	-0.0385	-0.155	0.115	0.47	0.238	0.945
Value of Reliability	-0.257	-1.13	-0.189	-1.22	-0.115	-0.985	-0.0448	-0.60	-0.160	-1.69
N	103		81		82		84		350	
Mean LL per obs	-0.610		-0.843		-0.833		-0.839		-0.818	
adj. ρ^2	0.62		0.47		0.48		0.48		0.53	

In the Monte Carlo experiments in Section 4, the true values for all parameters were known to us, and this is of course no longer the case here. Thus, in this case we take the RP results as the benchmarks, assuming they do not suffer from endogeneity. The aim of adding SP data to RP data in such a context is of course then to not change the parameters away from the RP estimators (which are consistent) but to improve efficiency by adding additional observations and/or to facilitate empirical identification by adding variance to the model attributes.

For the RP data, we see a negative estimate for rate and time, where neither is statistically significant at usual levels of confidence, along with a positive estimate for reliability. Because of the small sample size, we see numerous problems with significance in the separate models for the three SP datasets. The jointly estimated RP/SP model performs better, albeit that the coefficients for rate and time are still not significantly different from zero at usual levels of confidence. The model also indicates a lower scale for the SP data, with $\mu_{SP} < 1$. The findings on the pooled RP/SP data show that changes in the estimates (compared to RP) are not consistent across parameters, but they overall remain in line with

---

[6] Estimated using the "sandwich" estimator of the covariance (see e.g. Train, 2009, Section 8.6)

the RP estimates (which are consistent), while, across all three marginal utility parameters, we see increases in significance, which is the main benefit of adding the SP data.

We next proceed with a discussion of the results using the FIML correction[7], where we add five models beyond those reported by T&W. These models all estimate ρ, following van Cranenburgh et al. (2014), where we do this for both the overall sample and for three models combining the RP data with one of the SP datasets at a time. Finally, we include another model adding an error term to capture correlation across the SP choices due to a random effect (T&W$\rho\sigma$).

The results of these various models are reported in Table 3. We see that each of the three pooled models produces better fit than the pooled model in Table 2. This is in itself not surprising as the T&W models capture correlation across the individual choices. What remains to be answered is whether this is due to capturing endogeneity or other correlation, a point we look at in detail.

We begin by looking at the three models estimated using data from only one of the three SP experiments (alongside RP), but estimating the ρ error transfer parameter. With the FIML approach, the models are jointly estimated on the RP and SP data, and as a result, we see an increase in the number of observations per individual (and then in the sample size) compared with Table 2. We also see improvements in the per observation LL, a pattern repeated in the adjusted ρ^2 measure, reflecting the fact that the RP choices had better fit than the SP ones. For these individual models, the findings for the SP scale are in line with the results of the pooled RP/SP model in Table 2, i.e. showing lower scale for the SP data than the RP data, a finding that is also supported by a simple comparison of the estimates for the three separate SP models in Table 2 and the RP model. In the context of the present paper, the most important finding from these three separate models in Table 3 relates to the estimate for ρ, which is not significantly different from 0 in any of the models, thus not providing any evidence of error transfer from the RP to the SP data.

We next move to the original T&W model (column 5). This model produces a substantially higher value of time than the RP model, which would be a troubling finding since the RP model is supposed to produce consistent estimates. Additionally, and contradicting the findings from Table 2 as well as the SP-specific models, we see that the estimate for μ_{SP} is larger than 1, implying higher scale for SP than RP. Part of the issues are resolved by the model which additionally estimates ρ (column 6). We see a reduction in μ_{SP} which is no longer significantly different from 1. In addition, the value of time estimate moves in the direction of the RP data. The model leads to a further improvement in fit and suggests the presence of substantial error transfer from the RP data to the SP data (ρ is significantly different from 0 and larger than 1).

However, when moving to our final model, in which we include the additional σ term to capture SP-specific correlation across the SP choice tasks due to a random effect, we see that the earlier findings were influenced by the presence of such correlation, leading to misguided interpretation. The first observation is that this new specification obtains further improvements in fit, owing to the large and significant additional σ term which captures correlation across the SP scenarios alone. More importantly, this model brings the value of time estimate very close to the RP value, and shows that the error transfer parameter (ρ) is not significantly different from 0. In other words, there is correlation between the individual

---

SP tasks and a model not allowing for this could misattribute this to correlation between the RP and the SP data. This suggests that the data generation process behind this real data may be like Case 4 of the Monte Carlo experiments, where endogeneity was not (or maybe lightly) present, but the use of the T&W approach neglecting potential exogenous sources of correlation between SP tasks due to a random effect may suggest the existence of endogeneity in the data. In this final model, we see that the scale parameter μ_{SP} for the SP data is lower than in the previous two models, and the finding that it remains above 1 can be explained by the fact that a large share of the SP specific noise is now captured by the additional error term σ, where this also explains the finding of $\mu_{SP} > 1$ in the two previous models, given the added noise through ρ.

Table 3. T&W Endogeneity Correction Portland SP-off-RP Freight Data

	RP		RP & SP1 (Change rate) T&W		RP & SP2 (Change time) T&W		RP & SP3 (Change relia.) T&W		Pooled RP/SP T&W with $\rho=1$		Pooled RP/SP T&W with ρ		Pooled RP/SP T&W with ρ and σ	
	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t
β_{Rate}	-0.125	-1.37	-0.211	-0.895	-0.156	-1.38	-0.171	-1.40	-0.205	-2.45	-0.239	-3.09	-0.188	-1.91
β_{Time}	-0.0342	-1.09	-0.0373	-1.11	-0.0383	-1.15	-0.0388	-1.19	-0.145	-4.13	-0.140	-3.16	-0.0506	-1.71
β_{Relia}	0.0322	2.34	0.0336	2.78	0.0334	2.81	0.0345	2.78	0.028	3.66	0.0282	4.27	0.0238	2.02
ASC_1	-1.74	-2.63	-1.45	-1.77	-1.56	-2.24	-1.52	-2.30	-0.126	-0.305	-0.405	-0.785	-1.50	-3.15
ASC_2	0	-	0	-	0	-	0	-	0	-	0	-	0	-
ASC_3	1.08	2.30	0.826	1.67	0.889	1.75	0.861	1.72	1.027	2.20	1.61	3.03	0.617	1.05
ASC_4	-0.675	-2.03	-0.728	-2.11	-0.783	-2.23	-0.787	-2.44	-0.123	-0.375	-0.0408	-0.105	-0.858	-2.24
ASC_5	-0.456	-0.435	-0.884	-0.615	-0.519	-0.435	-0.519	-0.415	-0.930	-0.795	-0.837	-0.655	-0.802	-0.765
ASC_6	-0.596	-0.425	-1.26	-0.435	-0.907	-0.505	-1.037	-0.525	-0.929	-0.595	-1.07	-0.585	-1.18	-0.585
μ_{RP}	1.00	-	1.00	-	1.00	-	1.00	-	1.00	-	1.00	-	1.00	-
μ_{SP}			0.307	1.01	0.273	1.85	0.406	2.09	5.66	2.31	3.56	1.67	2.67	2.07
ρ			0.805	0.715	0.444	0.625	0.705	1.31	1	-	1.32	5.08	0.599	1.57
σ													3.44	3.04
Value of Time	0.273	0.875	0.176	0.695	0.245	0.905	0.228	0.875	0.710	2.85	0.585	2.10	0.269	1.17
Value of Reliability	-0.257	-1.13	-0.159	-0.905	-0.214	-1.20	-0.202	-1.38	-0.136	-4.32	-0.118	-4.11	-0.127	-2.21
N	103		184		185		187		350		350		350	
Mean LL per obs	-0.610		-0.749		-0.744		-0.741		-0.690		-0.665		-0.506	
adj. ρ^2	0.62		0.55		0.55		0.56		0.60		0.61		0.70	

Table 4 finally reports the estimation results for the models incorporating the CF correction proposed in this paper. As mentioned before, the CF was applied in this case using the MLCF approach considering the joint estimation of both stages in a single maximum likelihood problem. This approach facilitates the calculation of the standard errors and makes the best of a database that has very few observations and low levels of variance.

Since in the experiment only one attribute of the chosen alternative is changed for each of the three SP experiments, the correction for endogeneity with the CF method in this case only involves the estimation and the inclusion of the respective residual of the first stage into the model. This means, for example, that for SP1, which considered a change in the rate of the chosen alternative, the following normal density must be added to the likelihood of each observation

$$f_n\left(\delta_{i\_n}^{Rate} \mid \alpha^{Rate}; \sigma_{Rate}\right) = \frac{1}{\sqrt{2\pi}\sigma_{Rate}} e^{-\frac{\left(\delta_{i\_n}^{Rate}-0\right)^2}{2\sigma_{Rate}^2}} , \tag{12}$$

where the residual $\delta_{i\_n}^{Rate}$ is calculated as

$$\delta_{i\_n}^{Rate} = Rate_{i\_n}^{SP} - \alpha_0^{Rate} - \alpha_1^{Rate} Rate_{i\_n}^{RP} - \alpha_2^{Rate} Time_{i\_n}^{RP} - \alpha_3^{Rate} Relia_{i\_n}^{RP} . \qquad (13)$$

Finally, the residual $\delta_{i\_n}^{Rate}$ is added as an auxiliary variable to the SP alternative that was chosen in the RP setting, allowing the estimation of the parameters $\alpha_0^{Rate}, \alpha_1^{Rate}, \alpha_2^{Rate}, \alpha_3^{Rate}$ and σ_{Rate}, which are reported in Table 4. An equivalent procedure is followed with SP2, SP3 and the pooled data models.

We begin by three models estimated on data combining the RP observations with the observations from just one of the three SP games, where we see that the model fits are broadly in line with those from the FIML models in Table 3. We next turn to the coefficients of the residuals θ_{cost}, θ_{time} and θ_{relia}. As noted by Rivers and Vuong (1988), the significance of these coefficients in the CF correction can be used as a test for the existence of endogeneity, which in this case means a test for the transfer of the RP error to the SP experiment. For the first two SP experiments the coefficient of the residuals $\hat{\theta}_{cost}$ and $\hat{\theta}_{time}$ are significantly different from zero around the 85% level of confidence, where such lower values could be accepted in specification tests of this kind. In any case, this suggests that the endogeneity problem may not be too severe in this data and/or that it is masked in the poor variability and small sample size available in this case. θ_{relia} for SP3 is clearly not significant, a finding that can be attributed to an absence of endogeneity or to the poor quality of the data. These findings are in line with the low statistical significance of ρ in the three separate models in Table 3, and explain the similarity in fit with those models. Across all three models, the value of time measure is broadly in line with the RP result, in any case more so than the results for the FIML approach in Table 3.

The more interesting results arise when looking at the final two models in Table 4. When we compare the pooled RP-SP model with CF correction (penultimate model in Table 4) to either of the two simple T&W models in Table 3 (third to last and second last), we see much lower model fit for the CF models. However, the value of time measure in the CF model is much closer to the RP results than the ones obtained with the FIML approach. The CF models do still highlight the presence of some endogeneity, but this is much less pronounced than with the FIML method. Further evidence that these FIML results were not reliable comes from the final CF model which incorporates the additional σ parameter to capture correlation across the SP choices. We see that this model obtains a model fit that is just as good as the final FIML model in Table 3 while still highlighting the presence of some endogeneity in the SP1 game. The value of time is again in line with the RP results. This suggests that the fit advantages for the third to last and second to last FIML model over the CF model without the additional error term were a result of capturing correlation across SP choices, rather than any ability to capture endogeneity.

Let us now return to the three objectives set out at the start of this section. We have clearly seen that the proposed CF method can be applied to the same problem as used previously with the FIML approach of T&W. This can be done at reduced computational

cost, while producing results that are consistent and with no apparent drop in efficiency. Most importantly, the method seems to be more robust in terms of avoiding misattributing endogeneity to other sources of correlations that may exist between the various SP observations.

Table 4. CF Endogeneity Correction Portland SP-off-RP Freight Data

	RP		RP & SP1 (Change rate) CF		RP & SP2 (Change time) CF		RP & SP3 (Change relia.) CF		RP & Pooled RP/SP CF		RP & Pooled RP/SP CF, additional σ	
	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t	est.	rob.t
β_{Rate}	-0.125	-1.37	-0.181	-1.50	-0.1427	-1.38	-0.151	-1.40	-0.174	-1.67	-0.178	-2.50
β_{Time}	-0.0342	-1.09	-0.0388	-1.17	-0.0418	-1.23	-0.0410	-1.27	-0.0448	-1.42	-0.0440	-1.42
β_{Relia}	0.0322	2.34	0.0321	2.74	0.0321	2.68	0.0307	2.55	0.0265	2.47	0.0222	2.23
ASC_1	-1.74	-2.63	-1.47	-2.12	-1.58	-2.26	-1.558	-2.3	-1.06	-1.59	-1.57	-2.94
ASC_2	0	-	0	-	0	-	0	-	0	-	0	-
ASC_3	1.08	2.30	0.708	1.38	0.827	1.61	0.748	1.44	0.0148	0.0200	0.458	0.710
ASC_4	-0.675	-2.03	-0.790	-2.36	-0.839	-2.41	-0.869	-2.63	-1.04	-2.94	-0.931	-2.50
ASC_5	-0.456	-0.43	-0.918	-0.740	-0.537	-0.460	-0.492	-0.410	-0.860	-0.670	-0.821	-0.920
ASC_6	-0.596	-0.42	-0.989	-0.520	-0.815	-0.490	-0.829	-0.470	-0.949	-0.520	-1.08	-0.650
μ_{RP}	1	-	1	-	1	-	1	-	1	-	1	-
μ_{SP}			0.297	1.63	0.291	1.73	0.345	1.97	0.500	1.87	3.45	2.69
θ_{cost}			0.658	1.39					0.469	1.68	0.588	1.41
θ_{time}					1.91	1.29			1.11	1.52	1.10	1.27
θ_{relia}							-0.0415	-0.68	-0.0351	-0.840	-0.0630	-1.01
σ											3.63	2.53
α_0^{Rate}			-5.47	-3.50					-5.47	-3.58	-5.36	-3.12
α_1^{Rate}			1.55	20.3					1.55	21.03	1.47	15.3
α_2^{Rate}			0.0713	1.65					0.0682	1.57	0.0961	1.96
α_3^{Rate}			0.0284	2.14					0.0293	2.24	0.0403	2.87
σ_{Rate}			2.24	15.3					2.24	15.3	2.26	15.1
α_0^{Time}					-1.49	-2.56			-1.52	-2.63	-1.38	-2.22
α_1^{Time}					0.0400	1.41			0.0371	1.36	0.0227	0.880
α_2^{Time}					1.49	47.9			1.49	47.77	1.49	47.8
α_3^{Time}					0.800E-3	0.110			0.00190	0.290	0.00300	0.460
σ_{Time}					1.14	11.0			1.13	11.1	1.13	11.4
α_0^{Relia}							22.0	2.84	22.3	2.92	24.05	3.10
α_1^{Relia}							-0.723	-1.80	-0.737	-1.85	-0.896	-2.14
α_2^{Relia}							-0.141	-0.540	-0.134	-0.520	-0.0797	-0.280
α_3^{Relia}							0.504	6.32	0.502	6.31	0.509	6.50
σ_{Relia}							12.9	15.4	12.9	15.4	13.0	15.4
V. of Time	0.273	0.87	0.215	0.915	0.293	0.936	0.271	0.927	0.257	1.05	0.248	1.23
V. of Relia.	-0.257	-1.13	-0.178	-1.33	-0.225	-1.20	-0.203	-1.24	-0.152	-1.80	-0.125	-2.14
N	103		184		185		187		350		350	
Mean LL per obs	-0.610		-0.744		-0.7269		-0.7387		-0.8017		-0.4985	
adj. ρ^2	0.62		0.55		0.55		0.54		0.51		0.68	

6. CONCLUSIONS

Stated Preference (SP) data allows analysts to study choices in a controlled experimental environment and at low cost but suffers from cognitive dissonance (or hypothetical bias) between the stated and the actual choice behaviour. To lessen this drawback, it is common practice to build SP attributes and alternatives from observed Revealed Preference (RP) choices. For example, in SP-off-RP data, the chosen RP alternative is worsened, and other

23

alternatives are improved to induce a change in behaviour. This practice may increase the realism of the choice task, compared to SP settings that ignore the RP data, but with the undesired cost of introducing endogeneity. This source of endogeneity was largely neglected in the literature and in practice until Train and Wilson (2008) proposed a FIML solution that can be implemented with simulation.

In this article, we propose a LIML approach to address the SP-off-RP endogeneity problem by using a method that does not need simulation, can be applied with standard software and uses data that is already available for the stated problem. The proposed method is an application of the control-function (CF) method to correct for endogeneity in discrete choice models, using the RP attributes as instrumental variables.

We discuss the theoretical and practical advantages and disadvantages of the CF and T&W methods and illustrate them using Monte Carlo and real data. Results show that both methods can successfully address the endogeneity problems that arise from the use of SP-off-RP data when the only source of correlation in the data is across both RP and SP choices, i.e. if there is no random effect exogenous errors shared only among the SP choice tasks. When such SP specific error exists, the CF continues to work but the T&W approach fails to properly address the endogeneity problem and may even show the presence of endogeneity in the data when this is not the case. The reason behind the failure of the T&W approach lies in the fact that it is a FIML approach that uses an incorrect specification for the likelihood when some of the correlation is only across the SP choices. As shown by the Monte Carlo experiments and real data application, this problem can be solved, e.g., by adding an error component term to disentangle the endogenous and the exogenous part of the correlation between the choice tasks.

On the other hand, the proposed CF method is more robust to the modelling assumptions considered, succeeding in correcting the endogeneity problem for all the cases analysed and not causing further problems when endogeneity was not present. The CF method has the additional advantage of being significantly easier to implement and less onerous from a computational perspective. The caveat is, of course, that the when the T&W method works, it proved to be more efficient at least with the Monte Carlo data, which could matter when the data available is of low quality. All else being equal, the FIML approach will fit better as it captures correlation across choices, where this can be addressed by adding an error component to the CF method, of course partly compromising its simplicity. One additional observation that can be made is that the CF approach will of course work best in those cases where the RP data provides strong instruments for the SP data – this would imply using proper shifts from the RP data to ensure that the power of the instruments is guaranteed.

Train and Wilson (2008) offered a much-needed solution for addressing endogeneity in SP-off-RP data. This seminal work has been widely cited but rarely applied, arguably because of its various practical difficulties. As a result, SP-off-RP surveys have also failed to be widely applied. In this paper we try to help close this gap by proposing a CF method to achieve that goal, which we show to be robust (less sensitive) to the assumptions about the data generation process, effective in the correction of endogeneity and easy to implement with standard software. We think that this method may be an important step toward transforming this critical correction for endogeneity into a mainstream method and lead to wider use of SP-off-RP surveys.

ACKNOWLEDGEMENTS

REFERENCES

Allenby, G. Arora, N. and Ginter, J. (1995) Incorporating Prior Knowledge into the Analysis of Conjoint Studies. Journal of Marketing Research, 32 (2) 152-162 Published by: American Marketing Association.

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H., & Rao, V. (1994). Combining revealed and stated preferences data. Marketing Letters, 5(4), 335-349.

Bradley, M. A., & Daly, A. J. (1997). Estimation of logit choice models using mixed stated preference and revealed preference information. Understanding travel behaviour in an era of change, 209-232.

Guevara, C. A., & Ben-Akiva, M. E. (2012). Change of scale and forecasting with the control-function method in logit models. Transportation Science, 46(3), 425-437.

Guevara, C., & Ben-Akiva, M. (2006). Endogeneity in residential location choice models. Transportation Research Record: Journal of the Transportation Research Board, (1977), 60-66.

Guevara, C. A. (2005). Addressing endogeneity in residential location models Master Thesis, Massachusetts Institute of Technology.

Guevara, C. A. (2010). Endogeneity and sampling of alternatives in spatial choice models Doctoral dissertation, Massachusetts Institute of Technology.

Guevara, C. A. (2015). Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transportation Research Part A: Policy and Practice, 82, 240-254

Guevara, C. A. (2018). Overidentification tests for the exogeneity of instruments in discrete choice models. Transportation Research Part B: Methodological, 114, 241-253.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," Econometrica, 46, 931-959.

Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. Transportation Research Part B: Methodological, 44(6), 735-752.

Karaca-Mandic, P., & Train, K. (2003). Standard error correction in two-stage estimation with nested samples. The Econometrics Journal, 6(2), 401-407.

Lee, L. (1982), "Specification Error in Multinomial Logit Models," Journal of Econometrics, 20, 197–209.

Petrin, A. and K. Train (2002), "Omitted Product Attributes in Discrete Choice Models," Working Paper, Department of Economics, University of California, Berkeley, CA.

Rivers, D. and Q. Vuong (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," Journal of Econometrics, 39, 347-366.

Rose, J.M. & Bliemer, M. (2014) Stated choice experimental design theory: the who, the what and the why, in S. Hess & A. Daly, Handbook of Choice Modelling, chapter 7, pp. 152-177, Edward Elgar, Cheltenham.

Ruud, P. (1983), "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Models," Econometrica, 51, 225–228.

Shinghal, N., & Fowkes, T. (2002). Freight mode choice and adaptive stated preferences. Transportation Research Part E: Logistics and Transportation Review, 38(5), 367-378.

Terza, J. V. (2018). Two-Stage Residual Inclusion Estimation in Health Services Research and Health Economics. Health services research, 53(3), 1890-1899.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. Journal of health economics, 27(3), 531-543.

Train, K. E., & Wilson, W. W. (2009). Monte Carlo analysis of SP-off-RP data. Journal of Choice Modelling, 2(1), 101-117.

Train, K., & Wilson, W. W. (2008). Estimation on stated-preference experiments constructed from revealed-preference choices. Transportation Research Part B: Methodological, 42(3), 191-203.

Van Cranenburgh, S., Chorus, C. G., & Van Wee, B. (2014). Vacation behaviour under high travel cost conditions–A stated preference of revealed preference approach. Tourism Management, 43, 105-118.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.