Science of the Total Environment 633 (2018) 967–980

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A multiple indicator solution approach to endogeneity in discrete-choice models for environmental valuation

Petr Mariel [a,b,*], David Hoyos [a,b,c], Alaitz Artabe [b,d], C. Angelo Guevara [e]

[a] Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country (UPV/EHU), Spain
[b] Avda. Lehendakari Aguirre, 83. E48015 Bilbao, Spain
[c] EKOPOL, Research Group on Ecological Economics and Political Ecology, Spain
[d] Department of Applied Economics II (Public Finances and Tax Law), University of the Basque Country (UPV/EHU), Spain
[e] Departamento de Ingeniería Civil, Universidad de Chile, Casilla 228-3, Santiago, Chile

HIGHLIGHTS

- The paper applies the multiple indicator solution method in a discrete choice model for environmental valuation.
- It compares two different approaches to deal with endogeneity arising from omitted explanatory variables.
- The multiple indicator solution method and the hybrid model approach provide similar results in terms of welfare estimates.
- The multiple indicator solution method is more parsimonious and notably easier to implement but less efficient and flexible.

GRAPHICAL ABSTRACT

Environmental Valuation Studies

How shall we deal with endogeneity arising from omitted explanatory variables in discrete choice models for environmental valuation?

Hybrid Choice Model	Multiple Indicator Solution
Similar parameter estimates	
(-) Higher estimation cost	(+) Easier estimation
(-) Specific code or econometric software package needed	(+) Easily applicable in a general econometric software
(+) Higher efficiency of the estimation method	(-) Lower efficiency of the estimation method
(+) More flexible method. Latent variable can be incorporated to any part of the model.	(-) Attitudinal latent variable can be incorporated only at individual level (allocation function)

ARTICLE INFO

Article history:
Received 13 December 2017
Received in revised form 21 March 2018
Accepted 21 March 2018
Available online xxxx

Editor: F Frederic Coulon

Keywords:
Discrete choice
Valuation
Multiple indicator solution
Choice modelling
Endogeneity
Stated preferences

ABSTRACT

Endogeneity is an often neglected issue in empirical applications of discrete choice modelling despite its severe consequences in terms of inconsistent parameter estimation and biased welfare measures. This article analyses the performance of the multiple indicator solution method to deal with endogeneity arising from omitted explanatory variables in discrete choice models for environmental valuation. We also propose and illustrate a factor analysis procedure for the selection of the indicators in practice. Additionally, the performance of this method is compared with the recently proposed hybrid choice modelling framework. In an empirical application we find that the multiple indicator solution method and the hybrid model approach provide similar results in terms of welfare estimates, although the multiple indicator solution method is more parsimonious and notably easier to implement. The empirical results open a path to explore the performance of this method when endogeneity is thought to have a different cause or under a different set of indicators.

1. Introduction

Discrete choice experiments (DCEs) are increasingly being used to elicit preferences for environmental and natural resources. Many

* Corresponding author at: Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country (UPV/EHU), Spain.

E-mail addresses: petr.mariel@ehu.eus (P. Mariel), david.hoyos@ehu.eus (D. Hoyos), alaitz.artabe@ehu.eus (A. Artabe), crguevar@ing.uchile.cl (C.A. Guevara).

https://doi.org/10.1016/j.scitotenv.2018.03.254

methodological issues have been addressed in the development of DCEs in the field of environmental valuation, including optimal experimental design, econometric models, attribute non-attendance, and so forth. However, the issue of endogeneity and how to deal with it has received little attention (Hoyos, 2010). This apparent non-interest is especially surprising when most authors agree that endogeneity has the potential to distort inferences about preferences and the policy recommendations that could be derived from the analysis of the choice data (Louviere et al., 2005).

Endogeneity refers to the existence of correlation between the deterministic part and the error term of a regression or choice model. In the presence of endogeneity, parameter estimates are inconsistent thus invalidating any inference obtained from the model. Although this issue may be unavoidable in many practical cases, the severe consequences that the existence of endogeneity may cause, requires special attention to investigate how to deal with it in the field of DCEs for environmental valuation. The main source of endogeneity in environmental valuation may be found in the omission of contextual conditions in the choice situations, either due to the impossibility of the researcher to measure omitted attributes in the choice tasks, or to extract the exact measure that was inferred by the respondents when making their choices. For instance, we may find endogeneity when modelling environmental decisions due to the omission of latent environmental attitudes. When selecting an alternative, the decision maker may take into consideration his/her pro-environmental beliefs or attitudes. As a result, if endogeneity is not properly addressed, the estimated coefficients will likely be biased.

Dealing with endogeneity in classical regression models is well established in the econometric literature (see e.g. Bun and Harrison, 2014; Wooldridge, 2010). However, many aspects of how to deal with this problem in the framework of non-linear models, like discrete choice models, are still under development and have received scarce attention in areas such as transportation and environmental economics. The control function (CF) method has been found to properly address endogeneity in discrete choice models when this problem arises at the alternative level (Guevara and Ben-Akiva, 2006; Guevara, 2015; Petrin and Train, 2010). As with the case of linear models, one critical step when applying the CF method in discrete choice models is finding a valid instrument for each endogenous variable. A proper instrument needs to be correlated with the endogenous variable but, at the same time, uncorrelated with the error term. In practice, finding a valid instrument is sometimes problematic, thus motivating the search for alternative methods such as the multiple indicator solution (MIS).

The MIS procedure was originally proposed in the late 1960s by Blalock Jr. and Costner (1969), and Costner (1969), for sociological models. More recently, Wooldridge (2010) formalised the method for linear models, and Guevara and Polanco (2016) adapted it for DCEs. The MIS method requires a minimum of two indicators and is applied in two steps. In the first step, one of the indicators is added to the structural equation of the model as an additional explanatory endogenous variable. In the second step, the endogeneity is dealt with using the latter indicator as an instrumental variable for the former one. While Wooldridge (2010) uses the two-stage least squares (2SLS) method to address endogeneity in linear models, the MIS method has been extended to discrete choice models by using the CF method in the second stage (Guevara and Polanco, 2016).

In this paper, we present an exploratory analysis of the performance of the MIS method to deal with endogeneity arising from omitted attributes in a DCE for environmental valuation. To our knowledge, this is the first application of the MIS method to correct for endogeneity in the context of discrete-choice models for environmental valuation. The performance of this method is compared with another treatment of endogeneity recently used in environmental valuation, namely hybrid choice models.

Hybrid choice models allow for the incorporation of latent behavioural constructs within the traditional choice models, and therefore, they can be seen as a solution for endogeneity caused by the omission of a relevant variable. Hybrid choice models were first proposed by McFadden (1986) and Train et al. (1987) and have been increasingly used in the last decade. Despite some criticism (Chorus and Kroesen, 2014), their applications can be found in transportation (Paulssen et al., 2014; Bhat et al., 2015; Thorhauge et al., 2015), environment (Hess and Beharry-Borg, 2012; Hoyos et al., 2015) or health economics (Kløjgaard and Hess, 2014). We find that the MIS method and the hybrid model approach provide similar results in terms of model fit and parameter interpretation, but MIS is more parsimonious and notably easier to implement.

The rest of the paper is structured as follows: Section 2 addresses the methodological issues behind the MIS method to deal with endogeneity in discrete choice models; Section 3 describes the empirical study in which this methodology will be tested along with previous results; Section 4 provides the main results of the paper; and Section 5 finishes by discussing the main findings of this investigation and suggesting future lines of research in the area.

2. Methodology

We depart from a classical structural equation for a choice model given by the random utility theory, which is used to link the deterministic model with a statistical model of human behaviour. Under this framework, the utility of alternative i for respondent n is given by[1]:

$$\begin{aligned} u^*_{in} &= x'_{in}\beta + \beta_q q^*_{in} + e^*_{in} = x'_{in}\beta + \varepsilon^*_{in} = v_{in} + \varepsilon^*_{in}, \\ y_{in} &= 1\left[u^*_{in} \geq u^*_{jn};\ \forall j \in c_n\right], \end{aligned} \tag{1}$$

where, utility u^*_{in} is a latent variable that cannot be observed by the researcher but, instead, an indicator y_{in} is observed, which takes the value one if the utility of alternative i is the largest among those in the choice set c_n, and takes the value zero otherwise. The latent utility that an individual obtains from an alternative depends linearly, with coefficients β, on a set of explanatory variables or attributes collected in a row vector x_{in}, on a latent (omitted) variable q^*_{in} and on an error term e^*_{in}. In the set of explanatory variables x_{in} we may find a set of attributes of alternative i for respondent n, the first element of this vector being a one (for all but one alternative), accounting for an alternative specific constant. In the case where the omitted variable q^*_{in} is correlated with any variable included in x_{in}, there is an endogeneity problem in the model because ε^*_{in} is correlated with v_{in} in this case.

The CF approach is a common method used in linear regression analysis that can address endogeneity problems at the level of each alternative in discrete choice models. However, as with the case of linear regression, a critical requirement for applying this method is finding valid instruments for the endogenous variables. An instrument for an endogenous variable is valid if it is, at the same time, correlated with the endogenous variable and independent (not only uncorrelated as in the 2SLS method for linear models) of the error term of the model (Guevara and Polanco, 2016). The MIS method provides valid instruments that can then be used in the CF approach.

Let us assume that, instead of the latent variable q^*_{in}, the researcher observes two indicators q_1 and q_2 generated by the following equations:

$$\begin{aligned} q_{1in} &= \alpha_{10} + \alpha_{11} q^*_{in} + e^*_{q1in}, \\ q_{2in} &= \alpha_{20} + \alpha_{21} q^*_{in} + e^*_{q2in}, \end{aligned} \tag{2}$$

[1] For clarification purposes, we denote with an asterisk those variables that are latent in order to distinguish them from those that are observed by the researcher. Model coefficients are denoted with Greek letters.