Transportation Research Part B 114 (2018) 241–253

Contents lists available at ScienceDirect

Transportation Research Part B

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

Overidentification tests for the exogeneity of instruments in discrete choice models

C. Angelo Guevara

Departamento de Ingeniería Civil, Universidad de Chile, Facultad de Ingeniería y Ciencias Aplicadas, Blanco Encalada 2002, Santiago, Chile

ARTICLE INFO

Article history:
Received 17 March 2018
Revised 27 April 2018
Accepted 30 May 2018

Keywords:
Endogeneity
Overidentification
Control-function
Logit
Probit
Instrumental variables

ABSTRACT

Endogeneity is often present in discrete choice models, precluding the consistent estimation of the model parameters. To correct for this problem, the researcher needs to gather exogenous instrumental variables, which should be independent of the error term of the model. This critical assumption can be tested using overidentification tests that rely on having more instruments than endogenous variables. For discrete choice models, instruments' exogeneity can be assessed using the Amemiya-Lee-Newey test, which relies in the estimation of an auxiliary GMM model build from reduced-form estimates. This paper proposes two alternative tests that are constructed as adaptations of the Refutability test and the Hausman test, into the discrete choice framework. The Refutability test consists in including instruments as additional variables in an auxiliary model that was corrected for endogeneity using all instruments available. The Hausman test is built from the comparison of the estimates attained using different subsets of instruments. Using a binary choice Monte Carlo experiment, the three tests are assessed in terms of size, power, and robustness to De Blander's condition for which all overidentification tests of this kind are blind. Results show that what is termed the modified Refutability test, which includes all instruments simultaneously, has smaller size distortion, larger power, and more robustness, compared to the state of the art Amemiya-Lee-Newey test and to the proposed Hausman test. Besides, the Amemiya-Lee-Newey and both versions of the Refutability test allow being agnostic about which instrument might be exogenous. The paper finishes highlighting the methodological and practical implications and limitations of these findings and suggesting future lines of research in this area.

1. Introduction

When the error term of a discrete choice model is not independent of the observed variables, conventional estimators of the model parameters are inconsistent, making the model misleading for behavioral assessment and policy design (see e.g. Guevara and Thomas, 2007). This problem is known as endogeneity and may be caused by three main reasons: errors in variables, simultaneous determination and the omission of attributes of alternatives (Guevara, 2015). Endogeneity is common to several types of discrete choice models used in transportation analysis, including, but not limited to the following: airline itinerary choice (Lurkin et al., 2017), mode choice (Fernandez-Antolin et al, 2016), passenger booking timing (Wen and Chen, 2017), learning models of route choice (Guevara et al., 2017), mobility data collection (Zegras et al., 2018),

E-mail address: crguevar@ing.uchile.cl

https://doi.org/10.1016/j.trb.2018.05.020
0191-2615/

demand for electric vehicles (Helveston, 2016), vehicles' purchases (Petrin and Train, 2010), valuation of public transport attributes (Guevara et al., 2018) and residential choice (Guevara 2005, 2010; Guevara and Ben-Akiva 2006, 2012; Ferreira, 2010; Guevara and Polanco, 2016).

The canonical methods to correct for endogeneity in discrete choice models rely on the availability of suitable instrumental variables that must comply with two seemingly opposing properties: relevance and exogeneity. On the one hand instruments must be correlated with the endogenous variable (relevant) and on the other they must be independent of the error term of the model (exogenous). Therefore, finding suitable instruments in practice can be challenging and even polemic (see e.g. Hausman, 1997; Bresnahan, 1997).

Mumbower et al. (2014) reviewed different types of possible sources for instrumental variables that have been proposed in the literature, classifying them into four categories. The first category (cost-shifting) corresponds to instruments that are based on variables that share marginal costs with the endogenous variable, but that differ in demand shocks that explain the error term. The second category (Stern-type) uses instruments that are built from measurable differences in market power positions among markets. The third category (Hausman-type) takes advantage of differences in geographical contexts to achieve the cost-shifting goal. The final category (BLP-type) corresponds to instruments that are based on measures of non-price characteristics of other products supplied by the same firm.

For example, Guevara and Ben-Akiva (2006, 2012) use what can be classified as Hausman type instruments to address a price endogeneity problem in a residential location choice model. The source of endogeneity in this case was that dwelling price was likely correlated with omitted attributes. The problem was addressed by the authors using the prices of other dwellings located not too far but neither too close to the incumbent dwelling. By this, relevance would be achieved because close dwellings may share marginal cost, and exogeneity would hold because dwellings are located beyond some limit that permits assuming that they differ in the demand shocks that may partially explain the error term of the model.

In practice, instrumental variables must be gathered based on the researcher's judgment on the validity of the proposed source for cost-shifting, Stern, Hausman or BLP instruments. However, without formal tests, the suitability of the proposed instruments is always arguable. Testing if the instruments are relevant could be achieved by analyzing the correlation between the instrument and the endogenous variable, which are both observable. The challenge in such a case is to determine a statistic that can guarantee, with certain confidence, that the instruments are sufficiently correlated with the endogenous variable to attain certain goal in bias or empirical coverage (Guevara and Navarro, 2015). Testing instruments exogeneity is somehow more difficult because the error term of the model is latent (not observed), hindering the construction of a statistic to judge their independence from the instrumental variables. This article will tackle the latter challenge in the context of discrete choice models.

The challenge of judging the independence between the instrumental variable and the latent error for assessing exogeneity, is achieved by relying on overidentification. To correct for endogeneity, only one instrument per endogenous variable is needed. If more than the minimum needed instruments are used, the corrected model is said to be overidentified. If all instruments are exogenous, estimators will be consistent, and then, e.g., any difference attained with alternative sets of instruments used could only be attributed to sampling error, which allows testing instruments' validity. For linear models, Sargan (1958) noted that when the problem is overidentified the residuals of the instrumental-variables regression can be used to test for the exogeneity of the instruments. For nonlinear models, including discrete choice models such as the logit or the probit, Lee (1992) noted that an estimator developed by Amemiya (1978), and studied by Newey (1987), can play the role of the Sargan test. This test is usually termed as the Amemiya-Lee-Newey (ALN) test, it relies on the estimation of an auxiliary GMM model build from reduced-form estimates, and it is the state of the art in the subject.

Overidentification tests for the exogeneity of the instruments have a key limitation. Newey (1985) showed that these tests are inconsistent, which means that they are blind to certain alternative hypothesis even if the sample size goes to infinity. One way to recover consistency is to consider that the overidentification tests work under the assumption that a subset of the instruments, for which the model becomes just identified, is exogenous (Stock, 2001). This additional assumption cannot be proven, what seems to discourage the use of methods to correct for endogeneity that are based on instrumental variables.

However, De Blander (2008) proposes an alternative way to attain consistency of overidentification tests, which seems to increase their practical appealing. He notes that the alternative hypothesis for which overidentification tests are blind is very peculiar, so he recommends instead to assume that the conditions that produce this alternative hypothesis do not hold. This change puts "the burden of proof … on the critic, who has to make the case why the instruments" would fulfill this rare condition. De Blander (2008) shows that consistency would fail if the way in which the instruments appear in the structural equation and the reduced form equation, are linearly dependent. Pleus (2015) provides a more general expression for this result, building on Newey (1985), and presents a graphical representation to illustrate the nature of the problem. Parente and Silva (2012) identify one plausible case in which this peculiar condition may occur in practice, which is when both instruments are of the same nature, if they come from the same source. The problem is that, in such a case, the correlation of the instruments with both the endogenous variable and the error term will likely be very similar. An analogous warning was previously recommended by Nichols (2007), although justified on a different ground. Beyond Parente and Silva's (2012) warning, which is fully addressable in practice, it seems then easier to defend than to attack the plausibility of the consistency of the overidentification tests for the validity of instruments.

This article focuses on the development and the assessment of tests for the exogeneity of instruments into the discrete choice modeling framework. First, the state of the art Amemiya-Lee-Newey test is reviewed and compared with two novel