ASSESSMENT AND CORRECTION OF ENDOGENEITY PROBLEMS IN DISCRETE CHOICE MODELS

Thesis submitted to Newcastle University and the Pontificia Universidad Católica de Chile in partial fulfilment of the requirements to receive a dual Degree of Doctor of Philosophy and Doctor in Engineering Sciences

By Thomas E Guerrero

Supervisors:
Prof. Elisabetta Cherchi
Prof. Juan de Dios Ortúzar
Prof. C Ángelo Guevara

Newcastle University

Engineering School

August/2021

"[1] I waited patiently for the Lord; he inclined to me and heard my cry. [2] He drew me up from the pit of destruction, out of the miry bog, and set my feet upon a rock, making my steps secure. [3] He put a new song in my mouth, a song of praise to our God. Many will see and fear and put their trust in the Lord." (Salm 40)

ACKNOWLEDGMENTS

To God, for the infinite grace and blessings I have received. Even though I do not deserve anything I have and am, God has given me everything.

To my mother, my role model. Words are not enough to express my gratitude for the blessing you have been in my life. Thank you, mom!

To my siblings, Marcia and Camilo, my respect and admiration for you. I have learned from you the true love of brothers. Please, never change.

To Yanit and Miguel, Marilú, Nubia, Libardo, my aunts Marlene and Nubia, and my family in general, because they have taught me that even in the worst of times the family is forever, they are the ones that never fail. Always present, staunch as oaks.

To Juan de Dios (profe), because I never met someone who trusted me as much as you do. My respect and affection towards you will be forever. Pure wisdom, friend and advisor. Thank you, profe.

To Angelo, because apart from being my supervisor I was able to learn humility and nobility together in the same person. Thank you for sharing with me your valuable research ideas and knowledge. Thank you for all the time that you dedicated to me. The man behind the power!

To Elisabetta (Beba), I admire that beautiful smile that illuminates your face. Do not let it ever go away. Thank you for welcoming me to Newcastle University as your Ph.D. student. Your sympathy and affection remain deep in my heart.

To the Pontificia Universidad Católica de Chile (PUC), the *alma mater* that welcomed me for this Ph.D. I had not received this level of education before. Thank you for letting me develop as a student, as a professional, and as a person.

To the lectures, administrative staff and students of the Department of Transport Engineering and Logistics at PUC. Thank you, everybody.

To the Universidad Francisco de Paula Santander Ocaña, my second home in Colombia. My job. Thank you for the support received during this stage of my life.

To Newcastle University in the UK, the home that hosted me for a year. One of the most helpful and happy years of my life.

To all the amazing people that I met, or met again, during my doctorate. My sincere thanks to Paul Basnak (and his wonderful wife Pao), Margareth Gutierrez (and her invaluable husband Nelson), to my office partners (Yerly Martinez and Ignacio “Nacho” Tiznado) at Pontificia Universidad Católica de Chile, to the great Mitsuyoshi “Mitsu” Fukushi and to my Chinese friend Hao Yin (Andrew) at Newcastle University. Special mention to Julián Arellana, who advised me on the experimental design of the survey. Sorry if I miss anyone.

To all the researchers who preceded me and contributed to my own research.

Last, but not least, I would like to thank Chile and the Chilean people. I consider this country my second homeland. From this country and its people, I received more than I expected. I have a deep affection in my heart for the many Chileans I have met. If one day I can do something for this nation, I will do it with the greatest pleasure.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

NEWCASTLE UNIVERSITY

SCHOOL OF ENGINEERING

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

ASSESSMENT AND CORRECTION OF ENDOGENEITY PROBLEMS IN DISCRETE CHOICE MODELS

Thesis submitted to Newcastle University and the Pontificia Universidad Católica de Chile in partial fulfilment of the requirements to receive a dual Degree of Doctor of Philosophy and Doctor in Engineering Sciences.

THOMAS EDISON GUERRERO BARBOSA

ABSTRACT

The term *endogeneity* is used when there is a correlation between one or more observed explanatory variables (independent variables) and the error term of an econometric model. Endogeneity is considered a practically inevitable phenomenon in econometric modelling, as there are many potential causes behind it: omitted variables, measurement or specification errors, simultaneous estimation and self-selection. The problem is that it may give rise to inconsistent parameter estimates, and if its effects are not considered when estimating a model, the analyst may come to wrong forecasts and conclusions.

Correcting for endogeneity has been widely addressed in the linear models (LM) literature, but LM have a limited scope in certain areas. This is particularly the case in planning and social evaluation of transport projects, where Discrete Choice Models (DCM), which are highly non-linear, play a fundamental role. Unfortunately, DCM are

not often corrected for endogeneity, so a gap has been identified in the state of knowledge that this thesis intends to close. Thus, the general aim of this Ph.D. dissertation is to develop a set of guidelines that allow for the assessment and correction of endogeneity problems in DCM.

We establish conclusions of a theoretical, empirical and methodological nature. In the first instance, it is desired to determine adequate instrumental variables for endogeneity correction in transport modelling and measure the impact of this correction on strategic modal split models. We can reduce the errors associated with the estimation of DCM, improve its forecasting capabilities, and achieve consistent parameters resulting in corrected estimates of model valuation measures, such as the subjective value of time (SVT). Furthermore, we formulate an empirical methodology, supported by Monte Carlo simulation, to predict using DCM corrected for endogeneity with a new and more adequate version of the CF method. We also define guidelines to clarify under what conditions discrete indicators work (or not) when DCM are corrected for endogeneity using the MIS method. Finally, we structure a methodology to detect weak DCM instruments based on what has been proposed for linear models.

Newcastle, 2021

1. Introduction

1.1. Background

In econometric terminology, endogeneity is referred to an anomaly that yields inconsistent parameters. The problem behind is that if decision-makers support their analysis in endogenous models, these may be biased and wrong. Usually, endogeneity arises when there are a variety of conditions in the modelling, such as omitted attributes, measurement or specification errors, simultaneous determination and/or self-selection (Guevara, 2015).

Initially, the endogeneity problem was addressed in depth in linear models because of their wide general applicability. For example, some practical cases studied the consumer demand-supply analysis involving simultaneously interrelated variables (Wooldridge, 2010, Chapter 9). However, not all phenomena in science can be modelled using linear models. The cornerstone in the transport field is the classic four-stage transport model, where demand and supply interact (Ortúzar and Willumsen, 2011). In particular, for the third stage (known as *modal split* or *mode choice*) of the classic transport model, Discrete Choice Models (DCM) are typically used (Green, 1999; Ortúzar and Willumsen, 2011). The reason is that DCM are ideal when the dependent variable belongs to a finite set of discrete options. These models can be easily affected by endogeneity problems because, in modelling transport, it is common to find omitted attributes, measurement or specification errors, simultaneous determination and/or self-selection. If the DCM is not properly corrected for the effect of endogeneity, estimates may be biased, and the analysis may lead to potentially wrong decision making (Guevara and Ben-Akiva, 2006).

Several methods have been proposed to correct the problem of endogeneity in econometric models. A literature review is included in the next section. Although corrections for endogeneity using these methods are scarce in the transport field; some methods have been proposed to correct for endogeneity in DCM.

1.2. Literature Review and Research Gaps

As mentioned before, different methods have been proposed to correct for endogeneity. In the case of linear models, for example, the most used methods are the approach of Berry-Levinsohn-Pakes - BLP (Berry *et al*., 1995), Proxies (Wooldridge, 2010), Control Functions - CF (Heckman, 1978, Rivers and Vuong, 1988), and Multiple Indicator Solutions - MIS (Wooldridge, 2010). On the other hand, other approaches, or modified versions for linear models, have been also proposed in the literature for DCM, such as Latent Variables - LV (Ben-Akiva *et al*., 2002), Maximum Likelihood (Park and Gupta, 2009), CF (Petrin and Train, 2010), Proxies (Guevara, 2015), and MIS (Guevara and Polanco, 2016). An extensive compendium of advances in the treatment of endogeneity in DCM was developed by Louviere *et al*. (2005).

Given the objectives of this research, this review focuses only on methods proposed for DCM, and among them, we discuss two approaches in detail: CF and MIS. The reason is that these are the methods where research gaps have been detected, and they will be specifically studied in this Ph.D. thesis.

1.2.1. *Control Function (CF) method*

The idea behind the CF method is to build an auxiliary variable (or control function), such that when it is added to the systematic part of the DCM's utility function, it renders an explanatory variable considered endogenous appropriately exogenous (Barnow *et al*, 1981). The approach has been reported as a suitable approach for correcting endogeneity at the individual level (Train, 2009; Petrin and Train, 2010; Wooldridge, 2015). Besides the CF presents advantages due to its easy application and low consumption of computational resources (Guevara, 2015). The CF method can be estimated following the *Two Stage Least Square* - TSLS (Wooldridge, 2010) approach or simultaneously, in which case it is called *Maximum Likelihood* approach (Train, 2009).

Hausman (1978) showed that to apply the CF method, it is necessary to identify or construct a set of *Instrumental Variables* (IV, also known as *instruments*). Later, Rivers and Vuong (1988) and Villas-Boas and Winer (1999) proved that, formally, the IV are valid if they fulfil two properties: *(i)* be correlated with the endogenous variable, and *(ii)* be uncorrelated with the DCM's error. However, Bresnahan (1997) highlights that, in practice, identifying proper IV is always a difficult and even controversial process. In particular, the CF method can be hard to apply in the case of strategic transport modelling, because it is difficult and unclear how to obtain proper IV to correct the intrinsic endogeneity in these models.

In the DCM realm, the CF approach has been used by Mumbower *et al* (2014), Wen and Chen (2017) and Lurkin *et al* (2017). However, as Guevara (2015) highlighted, an important methodological challenge is associated with the practical difficulty of finding valid IV to correct for endogeneity. Noteworthy, we did not find any urban mode choice model included in a classic strategic transport model suite that has been corrected for endogeneity. Neither did we find valid IV that allow correcting for endogeneity in this type of models. Having detected these gaps in the literature, we intend to contribute to the state of knowledge by determining the impact of endogeneity in mode choice models estimated at the strategic level, by means of the CF approach with valid IV.

Several types of IV have been reported in the scientific literature; in certain cases, they have been found appropriate depending on the modelling context. A discussion on the theoretical soundness of IV as tools for correcting endogeneity is made by Bresnahan (1997). According to our review of the literature, adequate instrumental variables have been found for the case of airline choice (Granados *et al*., 2012; Hsiao, 2008; Berry and Jia, 2010; Gayle, 2004; Mumbower *et al*., 2014; Wen and Chen, 2017; Lurkin *et al*., 2017). However, we found no papers about using IV for correcting endogeneity in urban mode choice models. Since airline choice and urban mode choice involve rather different decision processes, we suspect that although there are IV functions proposed for the airline choice case, these cannot be used for strategic urban mode choice models.

1.2.2. *Multiple Indicator Solution (MIS) method*

Another approach to correct for endogeneity in econometric models and particularly in DCM, is the MIS method. Contrary to the CF method, the MIS uses indicators (instead of IV) to correct the model and achieve consistent estimators. Guevara (2015) shows some advantages of the MIS method, such as its easy applicability in practice. In several situations, the indicators may be easier to obtain given the difficulties already discussed regarding IV in the CF method.

In fact, while the IV must fulfil the two requirements described above, the indicators have a different nature. This is because, in practice, they come from questions in a survey designed for knowing the attitudes and/or perceptions that respondents have regarding their decision making. In the literature, indicators have been collected in several ways, such as Likert (1932) scales or verbal scales (Glerum *et al*., 2014). Besides, they can be *attitudinal* (Walker and Ben-Akiva, 2002; Daly *et al*., 2012; Bahamonde-Birke *et al.*, 2017) or *perceptual* (Bolduc and Daziano, 2009; Yáñez *et al*., 2010; Raveau *et al*., 2010).

Theoretically, to apply the MIS method to correct for endogeneity in DCM, the indicators must be continuous (Guevara, 2015) because the method is derived mathematically only for this case (Wooldridge, 2010). However, in practice the indicators tend to be discrete since they are typically obtained through Likert scales, and although empirical evidence suggests that discrete indicators may be as good as continuous ones for correcting endogeneity with the MIS method in DCM (Guevara, 2015; Guevara and Polanco, 2016; Fernández-Antolín *et al*., 2016), this can only be an approximation (Guevara, 2015). Given that there is not an answer for this in the literature, part of this thesis will focus on characterizing the impact of discrete indicators to correct for endogeneity in DCM using the MIS method.

1.2.3. *Weakness/strength of instruments*

A relevant concept related to the use of instruments to correct for endogeneity, corresponds to their *weakness*/*strength*. An IV can be considered *strong* when it is highly correlated with the endogenous variable; this allows estimating consistent parameters. On the other hand, an instrument is *weak* when there is a poor/slight correlation with the endogenous variable. This is an undesirable situation and may lead to biased estimation of parameters in the model (Stock and Yogo, 2005). In fact, correcting endogenous models with weak instruments may yield models with even worse performance than the uncorrected ones (Staiger and Stock, 1997).

The estimation of inconsistent parameters due to the weakness of the instruments was, for a long time, a problem largely ignored in econometric modelling. Staiger and Stock (1997) established the first advances that allow to determine the impact of weak instruments in the correction of endogeneity for linear models. They also addressed a qualitative description of practical recommendations to detect the weakness of the instruments. Finally, Stock and Yogo (2005) derived formal statistical tests to establish criteria that allow to differentiate between weak and strong instruments in linear models.

As can be seen, the identification of weak and strong instruments for the correction of endogeneity has been solved for linear models (Stock and Yogo, 2005). However, it has not been extended for DCM. This is relevant because DCM play an important role in modelling transport demand (Ortúzar and Willumsen, 2011). If a DCM is corrected for endogeneity using weak instruments, in addition to causing biased parameters, it could cause models with worse performance than those without correction. This is a research gap that will be addressed through this research proposal.

The methodology for linear models proposed by Stock and Yogo (2005), consists of tabulating a set of *critical values* that allow determining if an instrument can (or cannot) be considered weak. It is based on two criteria: *Relative Bias* (RB) and *Distortion Size* (DS) of the Wald (1943) test. The critical values tabulated by Stock and Yogo (2005) depend on the estimator of the IV that the modeller is using, the number of instruments (K_Z), the number of endogenous regressors (n), and how much bias (5%, 10% or more) the modeller considers a tolerable value. For this part of the research, we propose an alternative empirical approach to detect weak instruments in linear models and DCM, using Monte Carlo simulation. For this, we extend and adapt the Monte Carlo methodology proposed by Guevara and Navarro (2015), which is based on the RB criteria proposed by Stock and Yogo (2005).

1.3. Hypotheses

The hypotheses of this thesis are articulated with the objectives discussed below:

H_1: Although strategic urban mode choice models are susceptible to endogeneity, this problem can be corrected using the CF method with valid instrumental variables.

H_2: If discrete indicators are used instead of continuous ones, under certain conditions the endogeneity correction using the MIS method may fail in DCM.

H_3: It is possible to differentiate between weak instruments and strong instruments when correcting for endogeneity in DCM using the CF method.

H_4: The critical values to detect weak instruments in linear models proposed by Stock and Yogo (2005), can be adequately adapted to the DCM case.

1.4. Objectives

This thesis aims to develop a set of guidelines to allow for the assessment and correction of endogeneity problems in DCM. Three specific objectives are proposed to achieve it:

O_1: Determine the impact of endogeneity in mode choice models estimated at the strategic level. This objective includes the following sub-objectives:

i) Use data from the 2014 Great Valparaiso Origin–Destination Survey to detect endogeneity in strategic urban mode choice models.
ii) Find valid IV to correct for endogeneity in urban discrete mode choice models.
iii) Estimate DCM with the residuals coming from the CF approach.
iv) Assess the impact of endogeneity in forecasting with strategic transport models corrected for endogeneity; show that this implies simulating beyond the base year.

O_2: Develop a methodology to detect weak instruments in DCM using the CF method and Monte Carlo experiments.

i) Estimate the effect of the weak instruments in DCM using the CF method and Monte Carlo experiments.
ii) Build tables of critical values that allow to differentiate between weak and strong instruments, based on the relative bias criterion.

O_3: Characterize the impact of discrete indicators to correct for endogeneity in discrete choice models.

i) Design a Monte Carlo experiment to allow characterizing the effects of correcting for endogeneity using the MIS method with discrete and continuous indicators.
ii) Develop an application with real data obtained from a stated preference survey to a sample of individuals for the context of departure time choice.

1.5. Methodology

The methodology proposed to achieve each objective is as follow:

Objective 1: Determine the impact of endogeneity in mode choice models estimated at the strategic level

Data Collection: This process considers simulated and real data. For the former, Monte Carlo generated data were used to make forecasts with mode choice models in a strategic supply-demand equilibration setting. We considered three sources of endogeneity in this case: (*i*) measurement error, (*ii*) omitted variables and (*iii*) the simultaneous estimation of key variables in a supply-demand equilibration mechanism. For the real data application, we used the 2014 Great Valparaiso Origin–Destination Survey in Chile.

Model Specification, Estimation and Evaluation: The first step was to determine endogeneity in the models to be estimated. In Monte Carlo simulations, endogeneity is guaranteed since the modeller has full control over the characteristics of the data generated (Williams and Ortúzar, 1982). In the case of real data, the presence of endogeneity was verified according to the indications given by Rivers and Vuong (1988).

Once the endogeneity correction was made using the CF method and proper instruments (in the case of the mode choice model for Gran Valparaiso), the aim was to quantify the correction's impacts. We compared and quantified these effects in terms of estimating model parameters and computing subjective values of time (SVT) for both the endogenous and corrected models, in the base year and future situations.

Objective 2: Develop a methodology to detect weak instruments in DCM using the CF method and Monte Carlo experiments

The challenge here was to extend the findings of Stock and Yogo (2005) and Skeels and Windmeijer (2018) for linear models, to DCM. Therefore, we wished to tabulate the critical values that allow determining if an instrument is weak or not, in the DCM case.

To determine the critical values in the case of DCM, an empirical approach based on the construction of a data bank from Monte Carlo simulations was used. Stock and Yogo (2005) determined critical values for linear models using analytic derivations. Afterwards, Skeels and Windmeijer (2018) extended the results of Stock and Yogo (2005) to include more variation in the number of instruments and in the degree of relative bias. We used these critical values as a reference and developed a methodology to detect weak instruments in DCM using the CF method and Monte Carlo experiments. Our findings were validated for the case of linear models and then extended to the case of DCM. The critical values depend on the number of instruments used for correcting and the level of relative bias that the analyst is willing to tolerate.

Objective 3: Characterize the impact of discrete indicators to correct for endogeneity in discrete choice models

In this part, data of a different nature were used again. First, Monte Carlo experiments were conducted where endogeneity was assumed to arise due to omitting a variable correlated with an observed one. Second, a stated preference (SP) survey, where socio-economic data about the respondents, the current trip and indicators for the chosen context were also collected. Following the recommendations of Ortúzar and Willumsen (2011) regarding the process of designing and collecting SP data, four parts were considered:

Definition of the study context: The context selected was departure time choice. To the best of our knowledge, nobody has suggested indicators to correct for endogeneity using the MIS method for this specific case. Thus, suggesting indicators for this modelling context is a methodological contribution.

Experimental design and building the questionnaire: We based our design on the work of Arellana *et al*. (2012), the recommendations of Zwerina *et al*. (2005) regarding the balance of levels and minimum overlap in a SP survey, and the work of Rose and Bliemer (2009) regarding efficient experimental designs to allow estimating reliable parameters that do not depend on large sample sizes. On the other hand, block designs were used (Ortúzar and Willumsen, 2011, Chapter 3) to reduce the number of hypothetical choice scenarios and decrease respondent burden (Caussade *et al*., 2005).

Data collection: The survey was conducted on-line using the Qualtrics software. Since the experiment involved actual respondents, there was a risk of not recruiting enough participants yielding a low sample size. Thereby, given the objectives of the thesis, we aimed for a convenience sample. Therefore, the estimated model results and analyses with the information coming from this survey, do not intend to create public policy or be used for actual planning applications. The purpose is to consider a practical choice case where endogeneity may arise.

Model estimation: We estimated several models: A benchmark model (including all variables that took part in the SP experiment), an Endogenous model (excluding some explanatory variables of the SP experiment) and a model corrected for endogeneity using the MIS method (both using discrete and continuous indicators). This allows to quantify and compare the performance of discrete indicators versus continuous ones.

In the case of simulated data, we tested several specifications of the utility function, and estimated DCM corrected with the MIS method. Tests included varying the distribution of the indicator, the sample size, and the discretization process.

1.6. Contents and Contributions

This thesis describes the most relevant findings of the PhD research organised in four chapters (Chapter 2 to Chapter 5). Each chapter has also been published in a paper, as briefly described in the next subsections.

The rest of the thesis is organised as follows. Chapter 2 addresses the problem of endogeneity in strategic urban mode choice models. Chapter 3 deals with the critical question of how to make forecasts with strategic urban mode choice models in the case of supply-demand equilibration settings. Chapter 4 proposes an alternative empirical approach to detect weak instruments in linear models and DCM, using Monte Carlo simulation. Chapter 5 deals with the characterisation of the impact of discrete indicators to correct for endogeneity using the MIS method in DCM. In the final section, we discuss the main conclusions and future research directions of this research.

Chapter 2: Addressing Endogeneity in Strategic Urban Mode Choice Models

Endogeneity is a potential anomaly in econometric models, which may cause inconsistent parameter estimates. Transport models are prone to this problem and applications that properly correct for it are scarce. This chapter focuses on how to address this issue in the case of strategic urban mode choice models (i.e., the third stage of classic strategic transport models), possibly the main tool for the assessment of costly transport projects. To address this problem, we propose and validate, for the first time, adequate instruments that may be obtained from data that is already available in this context. The proposed method is implemented using the Control Function approach, which we use to detect and correct for endogeneity in a case study in Valparaiso, Chile. The effects arising from the neglected endogeneity in this case study reflect on an overestimation between 26-49% of the subjective value of time and an underestimation of 33-75% of modal elasticities.

This chapter has already been published:

Guerrero, T.E., Guevara, C.A., Cherchi, E. and Ortúzar, J. de D. (2020). Addressing endogeneity in strategic urban mode choice models. *Transportation* https://doi.org/10.1007/s11116-020-10122-y

Contributors:

Thomas E. Guerrero: Design, analysis and interpretation of results, draft manuscript preparation.

C. Angelo Guevara and Juan de Dios Ortúzar: Study conception, analysis and interpretation of results, draft manuscript preparation.

Elisabetta Cherchi: Analysis and interpretation of results, draft manuscript preparation.

Chapter 3: Forecasting with Strategic Transport Models Corrected for Endogeneity

The correction of endogeneity is a problem in strategic transport modelling; the question remains on how to make appropriate forecasts in this case. We propose a variation of the classical *Control Function* (*CF*) method, called *Control Function Updated* (*CFU*), which considers updating the endogeneity correction using information from the future equilibria. The proposed method is assessed using Monte Carlo simulation for a strategic transport model affected by three endogeneity sources, examining the equilibrium results for various future scenarios. We compare the *CFU* method by doing nothing and with the classical *CF* approach. The forecasts are evaluated in terms of recovering the true (simulated) travel times and two indices of fit. Results show that the endogenous (do nothing) model produces large biases in simulated travel times and poor goodness-of-fit measures that steeply worsen with time in future scenarios. The corrected models perform much better and, in particular, the new *CFU* approach shows statistically significant improvements over the classical approach in all scenarios tested.

This chapter is under review as follows:

Guerrero, T.E., Guevara, C.A., Cherchi, E. and Ortúzar, J. de D. (2020). Forecasting with strategic transport models corrected for endogeneity. *Transportmetrica A: Transport Science* https://doi.org/10.1080/23249935.2021.1891154

Contributors

Thomas E. Guerrero: Study conception, design, analysis and interpretation of results, draft manuscript preparation.

C. Angelo Guevara, Elisabetta Cherchi and Juan de Dios Ortúzar: analysis and interpretation of results, draft manuscript preparation.

Chapter 4: A Monte Carlo Method to Detect Weak Instruments: Application to Linear and Discrete Choice Models

Endogeneity is a pervasive problem in econometrics that precludes the consistent estimation of model parameters. The correction of endogeneity requires strong instrumental variables, that is, variables which are sufficiently correlated with the endogenous variable. The challenge, in this case, lies in determining critical values for feasible statistics to judge whether an instrument is strong or weak, under given criteria. This has been profusely studied for linear models, but the extension of those results to discrete choice models is still incipient. In this chapter, we contribute to bridging this gap. For this, we propose a Monte Carlo method to identify weak instruments, which we successfully validate by contrasting its results with those reported by analytical procedures applied to linear models. Upon this validation, we are also able to recommend critical values for the single instrument problem in linear models, something that has been controversial and not fully solved yet. We then use the proposed Monte Carlo method in a discrete choice logit model, to test the hypothesis that the critical values based on the F-statistics of the first stage regression of the Control Function method, are the same as those reported for linear models. We also show that as in the case of linear models, the critical values depend on the number of instruments and how much bias, relative to the endogenous model, the modeller considers tolerable.

This chapter is under review as follows:

Guerrero, T.E., Guevara, C.A., Cherchi, E. and Ortúzar, J. de D. (2020). A Monte Carlo method to detect weak instruments: application to linear and discrete choice models. *Journal of Applied Econometrics* (under review).

Contributors

Thomas E. Guerrero: Study conception, design, analysis and interpretation of results, draft manuscript preparation.

C. Angelo Guevara, Elisabetta Cherchi and Juan de Dios Ortúzar: Study conception, analysis and interpretation of results, draft manuscript preparation.

Chapter 5: Characterizing the Impact of Discrete Indicators to Correct for Endogeneity in Discrete Choice Models

Endogeneity is a common problem in econometric modelling that may lead to estimating inconsistent parameters. In the scientific literature, the Multiple Indicator Solutions (MIS) method is used to correct for endogeneity. This approach uses indicators that, in practice, tend to be collected as discrete using Likert scales; however, theoretically, the MIS method is derived considering continuous indicators. To close this research gap, this paper focuses on characterizing the impact of discrete indicators when correcting for endogeneity using the MIS method in the case of discrete choice models. Our findings show that (*i*) under some conditions, using discrete indicators instead of continuous ones seems not to be a problem, however, (*ii*) there is also evidence that indicates that the correction could fail under not unusual circumstances.

This chapter is under review as follows:

Guerrero, T.E., Guevara, C.A., Cherchi, E. and Ortúzar, J. de D. (2021). Characterizing the Impact of Discrete Indicators to Correct for Endogeneity in Discrete Choice Models. *Journal of Choice Modelling* https://doi.org/10.1016/j.jocm.2021.100342

Contributors

Thomas E. Guerrero: Study conception, design, analysis and interpretation of results, draft manuscript preparation.

C. Angelo Guevara: Study conception, analysis and interpretation of results, draft manuscript preparation.

Elisabetta Cherchi and Juan de Dios Ortúzar: Analysis and interpretation of results, draft manuscript preparation.

2. Addressing Endogeneity in Strategic Urban Mode Choice Models

2.1. Introduction

In this chapter we intend to contribute to the state of knowledge by solving the following three challenges: (*i*) How is endogeneity detected in strategic urban mode choice models? This stage includes correcting the model; (*ii*) Solving the practical difficulty of finding adequate instrumental variables (IV) or instruments (Hausman, 1978) to correct for endogeneity in this context (Bresnahan, 1997; Guevara, 2010; Mumbower *et al*., 2014); the problem comes from the fact that the instruments must fulfil two conflicting properties: be correlated with the endogenous variable, and be independent of the model error; (*iii*) Quantifying the impacts of neglecting the problem of endogeneity in the estimation of strategic urban modal choice model's parameters.

The chapter is organised as follows. The theoretical framework section details the methodology used, with a focus on how endogeneity arises in DCM and the importance of defining appropriate instruments. The application section describes the databank used and its general characteristics; we present an endogenous model, its corrected version, the instruments used to correct it and a quantification of the impacts of using the corrected version. In the final section we discuss the main findings and conclusions of this part of the research.

2.2. Theoretical Framework

2.2.1. *Endogeneity and DCM*

DCM enjoy high applicability in econometrics (Train, 2009; Ortúzar and Willumsen, 2011). They are used when the dependent variable is discrete in the phenomenon studied, for example, when individuals must choose an alternative belonging to a finite set of options. The use of DCM is very common in areas such as transport demand (Yáñez *et al*., 2010; Bass *et al*., 2011; Jensen *et al*., 2013), road safety (Rizzi and Ortúzar, 2006; Anderson and Hernandez, 2017), marketing (Lam *et al*., 2010), spatial economics (Hurtubia and Bierlaire, 2014), tourism (Chou and Chen, 2014), urbanism (Torres *et al*., 2013) and environmental economics (Hess and Beharry-Borg, 2012).

Like other econometric models, DCM are not exempt from endogeneity, but methods, tests and effects differ from those observed in linear models. For example, the correction of endogeneity in DCM implies a change of scale in the estimated models, and this is not the case in linear models (Guevara and Ben-Akiva, 2012). While the problem has been addressed for many types of DCM, to the best of our knowledge it has not been studied under the framework of strategic urban mode choice models, suggesting a research gap that we want to fulfil with this research.

DCM are based on Random Utility Maximization (RUM), whereby the utility, U_{in} of a certain alternative *i* for an individual *n*, is explained by the analyst as the sum of an observed component (systematic, representative or measurable utility, V_{in}) and a random term (Domencich and McFadden 1975, Williams 1977), ε_{in} as shown in (2.1):

$$U_{in} = V_{in} + \varepsilon_{in} \tag{2.1}$$

here, V_{in} is a function of a set of observable and measurable attributes X_{ikn}, where the subscript *k* denotes the attribute; ε_{in} reflects individual tastes and idiosyncrasies not captured in X_{ikn}, in addition to any measurement errors or attributes omitted by the modeller.

This form allows explaining how two individuals with the same attributes and the same set of alternatives (*A*) available, can choose differently, or why an individual does not always select the best alternative (from the modeller's point of view, Ortúzar and Willumsen, 2011). Thus, individual *n* will choose alternative A_i belonging to her set of choices *A(n)* if and only if (2.2) is fulfilled:

$$U_{in} \geq U_{jn}, \forall A_i \in A(n) \tag{2.2}$$

If it is assumed that the errors follow an independent and homoscedastic (IID) Gumbel distribution (also called Extreme Value Type I), the popular Multinomial Logit (MNL) model is obtained (Domencich and McFadden, 1975); other assumptions about the nature and characteristics of the error term distribution will allow to define different models.

2.2.2. *The Control Function (CF) method*

It consists in identifying an auxiliary variable (or control function), such that when it is added to the systematic part of the DCM's utility function, it makes the error of the model uncorrelated with the observed variables (Guevara and Ben-Akiva, 2010). This auxiliary variable or CF is constructed by means of an instrumental variable (IV). The CF method has been used and reported as a suitable approach for correcting endogeneity (Train, 2009; Petrin and Train, 2010; Wooldridge 2015). Besides the application of the CF method for correcting endogeneity at the individual level presents the advantage of being easy to apply and requiring low consumption of computational resources (Guevara, 2015).

Rivers and Vuong (1988) and Villas-Boas and Winer (1999) among others, show that the IVs needed for the application of the CF method in DCM are valid if they fulfil two properties: (*i*) be correlated with the endogenous variable, and (*ii*) be independent of the DCM's error. However, identifying proper IVs in practice is always a difficult and even controversial process (see e.g., the debate in Bresnahan, 1997). In particular, the CF method can be hard to apply in the case of strategic urban transport modelling suites, because it is not clear how to obtain proper IV to correct for endogeneity in these models.

For explanatory purposes, we will consider a DCM with endogeneity due to the omission of a certain variable q. Assume that its true linear utility function is represented by (2.3):

$$U_{in} = ASC_i + \beta_x X_{in} + \beta_q q_{in} + e_{in} \qquad (2.3)$$

Where ASC_i is an alternative specific constant for alternative A_i; β_x and β_q are parameters to be estimated, X_{in} and q_{in} are explanatory variables of the model, and e_{in} is the exogenous error term. In particular, we will assume that X_{in} represents a set of known (measurable) attributes while the variable q_{in} is unknown to the modeller.

Given the above, let us assume that the specification proposed by the modeller is as in (2.4):

$$U_{in} = ASC_i + \beta_x X_{in} + \varepsilon_{in} \qquad (2.4)$$

where the new error term ε_{in} obviously contains both e_{in} and q_{in}. Now, let us consider that one of the elements of the set that makes up for X_{in} (for example, the *k-th* term) is correlated with q_{in}, as follows:

$$X_{kin} = \gamma_0 + \gamma_{z_1} z_{1in} + \gamma_{z_2} z_{2in} + \gamma_q q_{in} + \varphi_{in} \qquad (2.5)$$

where φ_{in} is an exogenous error term z_{1in} and z_{2in} are exogenous attributes, which then work as instruments or IV, since they partially explain X_{kin}, but are at the same time independent from ε_{in}. For the model to be identifiable, there must be (at least) as many IV as endogenous variables in the model (Guevara and Ben-Akiva 2012). Following the assumption that q_{in} is a variable not considered by the modeller, a specification that can be set up to treat potential endogeneity would be:

$$X_{kin} = \gamma_0 + \gamma_{z_1} z_{1in} + \gamma_{z_2} z_{2in} + \delta_{in} \qquad (2.6)$$

where the error term δ_{in} contains both φ_{in} and q_{in}. As it is now clear, endogeneity arises because the error terms ε_{in} (2.4) and δ_{in} (2.6) are correlated with each other, as q_{in} was not included in the model specification originally proposed by the modeller.

Thus, if (2.6) is valid, such that z_{1in} and z_{2in} are truly exogenous, then δ_{in} will capture the entire part of X_{kin} that is endogenous. This way, the DCM corrected by endogeneity using the CF approach would have the functional form shown in (2.7), which implies using a proper estimator of δ_{in}:

$$U_{in} = ASC_i + \hat{\beta}_x X_{in} + \beta_{\hat{\delta}} \hat{\delta}_{in} + \tilde{e}_{in} \qquad (2.7)$$

Thus, in practice the CF method follows two-stages:

(i) to obtain the residuals $\hat{\delta}_{in}$ by applying an ordinal least squares (OLS) regression to X_{kin} on z_{1in}, z_{2in} and all the exogenous variables in X_{in}.

(ii) to estimate the DCM considering $\hat{\delta}_{in}$ and the X_{in} attributes within the utility function.

This allows obtaining consistent estimators $\hat{\beta}_x$ for the β_x in (2.7) up to a scale (Guevara and Ben-Akiva, 2012), but the CF method can also be estimated simultaneously (Train, 2009). Theoretically, the two-stage estimation involves a loss of efficiency; however, as Rivers and Vuong (1988) show, this drawback may disappear when the error terms ε_{in} and δ_{in} in (2.4) and (2.6) are homoscedastic and not autocorrelated. The other drawback of the two-stage version of the CF method is that standard errors cannot be obtained directly from the information matrix, requiring alternative methods, such as the bootstrap. Nevertheless, as discussed by Guevara (2015), the two-stage version of the CF method is more robust to misspecifications of the distributional assumptions of the model, as well as much easier to apply and requiring fewer computational resources.

It worth noting that there are other methods to correct for endogeneity in DCM, beyond the CF method. These include, among others, the use of Proxies (Guevara, 2015), the Multiple Indicator Solution - MIS (Guevara and Polanco, 2016; Guevara *et al*., 2018; Mariel *et al*., 2018; Fernandez-Antolin *et al*., 2016), the latent variables approach (Walker, 2001) and the BLP method (Barry *et al*., 1995), among others. Guevara (2015) makes a critical assessment of most of these methods.

2.2.3. *Instrumental variables - IV*

A fundamental requirement that can turn into a real challenge for applying the CF method is the availability of proper IV. It is achieved if: (*i*) the IV are correlated with the endogenous variable, and (*ii*) the IV are independent of the DCM's error. The former is known as *relevance condition* and the second as *exogeneity condition*.

Mumbower *et al.* (2014) distinguish four possible sources for IV. The first are the *cost-shifting* instruments (Casey, 1989), which correspond to variables that impact a product's cost but are uncorrelated with demand shocks. The second are the so-called *Hausman instruments* (Hausman *et al*., 1994; Hausman, 1996), which correspond to prices of the same brand in other geographic contexts. The third are the *Stern instruments* used like measures of the level of market power by multiproduct firms and measures of the level of competition (Stern, 1996). Finally, the *BLP instruments* correspond to the average non-price characteristics of other products supplied by the same firm in the same market (Berry *et al.*, 1995). We will explore the potential of these types of instruments to correct for endogeneity in strategic urban mode choice models.

2.2.4. *Tests for the validity of instruments*

As mentioned above, a crucial challenge in the correction of endogeneity with the CF resides in finding proper instruments that are sufficiently correlated with the endogenous variable (*strong*) and independent of the error term (*exogenous*).

The strength of an instrument can be assessed by looking at the degree of correlation between the endogenous variable and the instrument, something that has been extensively investigated for linear models, but remains to be fully explored for DCM. Nevertheless, preliminary results suggest that this may be achieved looking at the F test of the first stage regression of the CF method, for which similar thresholds as those reported in linear models seem to be applicable (Guevara and Navarro, 2015).

Assessing the exogeneity of instruments is more challenging in some sense, because one needs to test independence with the error term, which is obviously not observed. This requirement may be guessed by the analyst based on his/her understanding of the data generation process but may also be formally tested with overidentification tests that rely on having more instruments than endogenous variables. In the case of linear models, the *Sargan* test (Sargan, 1958) is applicable. For DCM, the only test available until recently was the *Amemiya-Lee-Newey test* (Amemiya, 1978; Newey, 1987; Lee, 1992) that requires estimating an auxiliary generalized method of moments (GMM) model, making its application challenging. Guevara (2018) recently proposed two overidentification tests for the exogeneity of the instruments for DCM that are not only easier to apply, but also show better power and less size distortion[1] than the previous tests: The *Refutability Test* (S_{REF}) and its variation, the *Modified Refutability test* (S_{mREF}).

Guevara´s (2018) *Refutability Test* (S_{REF}) requires the following two stages:

Stage 1: Estimate the reduced form equation for X_{kin} in (2.6) by OLS to obtain the residuals $\hat{\delta}$, as shown in (2.8):

$$X_{kin} = \gamma_{z_1} z_{1in} + \gamma_{z_2} z_{2in} + \delta_{in} \xrightarrow{yields} \hat{\delta}_{in} \qquad (2.8)$$

Stage 2: Estimate the DCM considering $\hat{\delta}_{in}$ and the X_{in} attributes, but also one of the instruments (for example z_{1in}) as an additional variable within the utility function and obtain the log-likelihood $l(\theta)^{CF-Z1}$, consistent with the utility function shown in (2.9).

$$U_{in} = ASC_i + \beta'_x X_{in} + \beta_{\hat{\delta}} \hat{\delta}_{in} + \beta_{z_1} z_{1in} + \tilde{\tilde{e}}_{in} \qquad (2.9)$$

Given that in (2.9) only z_{1in} is used, a log-likelihood $l(\theta)^{CF\_Z1}$ is obtained. The same process must be repeated using z_{2in} as an additional variable within the utility function and obtain a log-likelihood $l(\theta)^{CF\_Z2}$. In this way, two log-likelihood values are computed, by fixing each time all instruments to zero but one (in our case by fixing z_{2in} first, and z_{1in} second).

The second test, S_{mREF}, can also be obtained in two stages. The first is the same as for the S_{REF} test (2.8); the second stage proceeds as follows:

[1] The size distortion corresponds to the difference between the nominal significance of the tests, and the empirical size for the Type I error under the null hypothesis. This type of measure is a standard tool for the assessment of the statistical tests (Guevara, 2018).

Stage 2: Estimate the DCM considering the ASC_i, β'_x and $\beta_{\hat{\delta}}$ fixed. Then add all the instruments considered (i.e., z_{1in} and z_{2in}) as additional variables within the utility function and obtain the log-likelihood $l(\theta)^{CF\_Zall}$, consistent with the utility function in (2.10):

$$U_{in} = ASC_i + \beta'_x X_{in} + \beta_{\hat{\delta}} \hat{\delta}_{in} + \beta_{z_1} z_{1in} + \beta_{z_2} z_{2in} + \tilde{\tilde{e}}_{in} \qquad (2.10)$$

The statistics of the *Refutability Test* - S_{REF} (2.11) and (2.12), and its modified version – S_{mREF} (2.13) – used to test for exogeneity are the following:

$$S_{REF}^{Fixing\ Z1} = -2(l(\theta)^{CF} - l(\theta)^{CF\_Z2}) \sim \chi_r^2 \qquad (2.11)$$

$$S_{REF}^{Fixing\ Z2} = -2(l(\theta)^{CF} - l(\theta)^{CF\_Z1}) \sim \chi_r^2 \qquad (2.12)$$

$$S_{mREF} = -2(l(\theta)^{CF} - l(\theta)^{CF\_Zall}) \sim \chi_r^2 \qquad (2.13)$$

where $l(\theta)^{CF}$ is the log-likelihood of the corrected model obtained in (2.7) and χ_r^2 is the value of the chi-squared distribution with degrees of freedom (r) equal to the degrees of overidentification of the model. For the reference tests described in (2.9) and (2.10), r is equal to 1 because the model includes one endogenous variable and two instruments (z_{1in} and z_{2in}). The null hypothesis for the S_{REF} and S_{mREF} tests is that both z_{1in} and z_{2in} are valid; the alternative hypothesis is that either z_{1in} and z_{2in}, or both, are endogenous. Thereby, if S_{REF} and S_{mREF} are less than the critical value of χ_r^2 at the required level of significance, the instruments are exogenous and, therefore, they are independent of the DCM's error.

It should be noted that overidentification tests for the exogeneity of the instruments are inconsistent, in the sense that there are null hypotheses for which the tests have no power. This means that there might be cases where the instruments are endogenous and these tests are unable to detect that failure, even if the sample size goes to infinity. Nevertheless, it has been shown that the hypotheses for which overidentification tests of this type are inconsistent, are very peculiar and can be narrowed to cases where both instruments are of the same origin, if they come from the same source. This is something we tried to avoid in the present application. The reader is referred to Guevara (2018, pp. 242) for a review and discussion about this topic.

2.2.5. *Subjective value of time (SVT) and elasticities*

We will estimate the SVT and aggregate elasticities to quantify the impacts of neglecting the problem of endogeneity in the estimation of the parameters of a strategic urban modal choice model. As the representative utility function in most classical models is assumed to be linear and additive in the (fixed) marginal utility parameters, the SVT (Gaudry *et al*. 1989) usually corresponds to just the ratio between the estimated parameters for travel time β_t and for travel cost β_c, yielding (2.14):

$$SVT = \frac{\partial V_i/\partial t_i}{\partial V_i/\partial c_i} = \frac{\beta_t}{\beta_c} \tag{2.14}$$

The aggregate elasticities ($E_{kin}^{\tilde{P}i}$) can be calculated as in (2.15):

$$E_{in}^{\tilde{P}i} = \frac{\sum_n P_n(i) E_{X_{in}}^{P_n(i)}}{\sum_n P_n(i)} \tag{2.15}$$

where $E_{X_{in}}^{P_n(i)}$ is the disaggregate direct point elasticity with respect to variable X_{in}, and $P_n(i)$ the probability that individual *n* chooses alternative *i* (Ben-Akiva and Lerman, 1985).

2.3. Application

2.3.1. *Great Valparaíso case study*

The Great Valparaíso is a conurbation located in the Valparaiso Region of Chile, encompassing the municipalities of Valparaíso, Viña del Mar, Concón, Quilpué and Villa Alemana, an area of some 1,130 km^2 (SECTRA 2014a). According to the National Statistics Institute (INE, 2013), it is the third most populated area in the country, after Great Santiago and Great Concepción, but given its strategic location and proximity to the capital, it is the second in importance.

The database comes from the Great Valparaiso 2014 Origin–Destination Survey and was used by SECTRA[2] (2014a) to estimate – among other things – the DCM embedded in the mode choice stage of ESTRAVAL, the strategic transport model for the Great Valparaiso. ESTRAVAL is a simultaneous supply-demand equilibrium model designed to analyse and evaluate multimodal urban transport systems with multiple user classes (De Cea *et al*. 2005). This type of approach is also used in packages such as EMME/2 (INRO, 1996) or CUBE (Citilabs, 2016).

The aim of our research was not to change the model used in ESTRAVAL; we just wanted to examine the consequences of correcting it for endogeneity. The model contemplates seven transport modes: Car driver, Shared car, Bus, Train, Shared taxi, Walking and the combined mode Train/Bus. The survey considered three trip purposes: Work, Study and Other, but in the framework of this research we only considered the correction of the work trips mode choice model for the morning peak period. The sample available for this purpose comprised 2,417 observations. A general descriptive statistical analysis of the sample shows that 36% of trips used private modes (Car driver and Shared car), while the most used public transport mode was Bus with 26% of the market shares.

[2] SECTRA is the Chilean governmental agency for transport planning and policy formulation.

2.3.2. *Instrumental variables used for endogeneity correction*

The instruments used to estimate the first stage of the CF method were built resembling what Mumbower *et al*. (2014) denominate *Hausman type* instruments, that is, values of the endogenous variable in "other markets", that may share marginal costs, but are independent regarding demand shocks. For this mode choice model, we suspect the existence of endogeneity both in travel time and travel cost, and for this reason we propose the following three instruments:

(i). The average travel time of other origin-destination (O-D) pairs with similar length to the O-D pair of the considered trip (*IV_GT*).
(ii). The average travel cost of other O-D pairs with similar length to the O-D pair of the considered trip (*IV_C*).
(iii). The network trip distance between the trip's origin and destination (*IV_D*).

As can be seen, each of these instruments should be correlated with the endogenous variables (cost, time or both) but they do not influence the individuals' choice, being then independent regarding demand shocks. If both properties are fulfilled, the instruments are valid to correct appropriately for endogeneity using the CF approach (Rivers and Vuong, 1988; Villas-Boas and Winer, 1999).

To verify that the proposed instruments fulfil the *relevance* condition, we considered the results described by Staiger and Stock (1997), which had been preliminary suggested to hold also for DCM by Guevara and Navarro (2015). In this case, if the value of the first stage's F-statistic is less than 10, the instrument is *weak* (i.e., it does not satisfactorily fulfil the condition). However, it should be noted that as this result formally holds only for linear models, this is a limitation of this research, which we intend to explore in the future.

To guarantee the exogeneity of the instruments, we considered using the information of a geographical context different from the Great Valparaiso (i.e., Hausman-type instruments), an approach which had been successfully used in several studies (Mumbower *et al*., 2014; Guevara and Ben-Akiva, 2006; Petrin and Train, 2010). In this case, the other geographical context data was the Santiago 2012 Origin–Destination Survey (SECTRA, 2014b) also in Chile. The procedure applied to find the instruments considered the zoning system used by SECTRA in their model for Santiago.

In this way, *IV_GT* and *IV_C* were calculated as the average travel time and travel cost, respectively, for every zone included in a band defined by a lower bound of ± 100 m and an upper bound of ± 2.1 km with respect to the distance of the O-D pair under consideration. For example, consider a distance (Euclidean distance, measured between centroids for the given O-D pair) of 5 km; in this case, the lower and upper bounds defined would be as follows: [2.9 km – 4.9 km] and [5.1 km – 7.1 km]. Thereby, any O-D pair, the distance of which is inside any of these two bands would be part of the average for *IV_GT* or *IV_C*. The argument to sustain the suitability of such instruments is equivalent to that used by Guevara and Ben-Akiva (2012), Hausman (1996) and Nevo (2001) in other modelling contexts.

The lower bound (100 m) guarantees that the O-D pair under consideration is not included because otherwise endogeneity would arise. On the other hand, the upper bound (2.1 km) ensures that every O-D pair has enough data to estimate an average. In this way, we make sure that every O-D pair has a set of O-D pairs inside the bands defined. This fact makes them share marginal travel costs (or travel times) and, therefore, their travel costs (or travel times) are correlated.

Finally, the third instrument used is the *IV_D*, known in the literature as a *cost-shifting* instrument (Casey, 1989). *IV_D* was calculated directly from the network defined for ESTRAVAL, thus, from the city of Valparaíso. Instruments of a similar nature (route distance) have also been used successfully for the case of air transport (Hsiao, 2008; Granados *et al*., 2012). We argue that *IV_D* is correlated with the travel time and travel cost, but independent of the error term of the mode choice.

It should be noted that any of the O-D pairs used to build the instruments (IV_GT and IV_C) could (or not) be overlapping among them. However, this is not an issue because the instruments were constructed as the attributes' average of the O-D pairs that were part of the bands defined above. What may instead be critical in general is that none of the O-D pairs used to build the instruments, overlapped with the O-D pair under analysis (i.e., the incumbent O-D pair for which we needed to address endogeneity). This is not necessarily an issue for IV_GT and IV_C in this case study, since they come from a different city, but it was nevertheless further enforced by defining the band's lower bound different from zero (100 m) to avoid endogeneity arising due to reflection bias. Regarding the IV_D instrument, the overlapping is also possible, but it did not affect the instrument estimation because it only depends on the route determined by the network topology used in ESTRAVAL.

2.3.3. *Correction of endogeneity in strategic urban mode choice models*

We assumed that endogeneity affects the travel cost and travel time variables in the Great Valparaíso urban mode choice model because of, as mentioned earlier, the potential erroneous measurement of the relevant variables included in the model, the omission of potentially relevant variables (such as comfort or reliability) and the fact that the model is embedded in a simultaneous supply-demand equilibrium mechanism. Our hypothesis is that the measurement error due to aggregation may affect both travel time and travel cost. And the omission of attributes and the simultaneity issue may affect travel time further.

Table 2-1 presents the endogenous and corrected mode choice models estimated with the Great Valparaiso dataset. The left-hand side model (potentially endogenous) is the model currently used by SECTRA in ESTRAVAL. This is the model that we want to correct for endogeneity. It was estimated by SECTRA for two morning peak periods (AM1 and AM2), so 14 (seven modes by two periods) alternative specific constants (ASC) were estimated, fixing one ($ASC_{Walking1}$) to zero, as reference. The parameter $\beta_{Cost/Income}$ corresponds to the marginal utility of the variable Cost divided by Income. The model also includes three different parameters for Generalised Time (i.e., the sum of travel time, access time and waiting time): $\beta_{Generalised\ Time\ Car}$, $\beta_{Generalised\ Time\ Walking}$ and $\beta_{Generalised\ Time\ Public\ transport,}$ correspond to the marginal utilities of the private modes (Car driver and Shared car), walk mode (Walking) and public transport modes (Bus, Train, Shared taxi and Train+Bus), respectively. Finally, $\beta_{Distance\ travel\ ST1}$, $\beta_{Distance\ travel\ TTB1}$, $\beta_{Distance\ travel\ ST2}$ and $\beta_{Distance\ travel\ TTB2}$ are parameters associated with dummy variables, which take the value of 1 for Shared taxi ($\beta_{Distance\ travel\ ST1}$ and $\beta_{Distance\ travel\ ST2}$ for the periods AM1 and AM2, respectively), and Train and Train+Bus ($\beta_{Distance\ travel\ TTB1}$ and $\beta_{Distance\ travel\ TTB2}$ for the same periods) if the trip had a distance greater than 10 km. All the level-of-service parameters of the potentially endogenous model in Table 2-1 have correct signs and are statistically significant at the 95% level. We note that trips with distances greater than 10 km are preferred by Train and Train+Bus users.

The right-hand side of Table 2-1 shows the model corrected for endogeneity. This includes the parameters $\beta_{\hat{\delta}GT}$ and $\beta_{\hat{\delta}CI}$ (residuals from the first stage of the CF approach) related to the variables Generalised Time and Cost/Income, respectively. The inclusion of these two parameters is required because of our initial hypothesis that the uncorrected model is endogenous in Cost/Income and in the Generalised Times. The verification of this hypothesis is carried out following Rivers and Vuong (1988); so, if $\beta_{\hat{\delta}GT}$ and $\beta_{\hat{\delta}CI}$ are significant in the second stage of the CF approach, then there is evidence that the model is endogenous in the variables related with these residuals. As can be seen from the right-hand side model, both $\beta_{\hat{\delta}GT}$ and $\beta_{\hat{\delta}CI}$ are significant.

One practical aspect of the application of the CF method to this case study[3], worth highlighting, is that the first stage of the CF method was estimated by mode, instead of stacking the information from all available alternatives, as has been done in other cases. This approach was followed because the Shared Car and Walking alternatives have a travel cost of zero in this application, and this would preclude proper estimation of the residuals of a stacked first stage via an OLS if the dependent variable is zero. Nevertheless, the same coefficient for the residual was considered for all modes, as shown in Table 2-1. An extensive Monte Carlo simulation validated this approach for the practical problem of modes with zero cost.

[3] An additional issue that did not come out in this application, but may be relevant for other cases, is what to do when the endogenous variable interacts with exogenous variables, such as level of income or gender. Bun and Harrison (2018) formally show that, under such circumstances, the endogeneity bias will reduce to zero for the ordinary least squares' estimator, as far as the interaction term is concerned. The same holds for the Control Function method in discrete choices, something that has been implicitly used, among others, by Petrin and Train (2010) and Guevara and Ben-Akiva (2006).

Variable	Endogenous model			Corrected model		
	Value	Std. error	t-test	Value	Std. error[a]	t-test
$ASC_{Car\ driver1}$	-0.425	0.329	-1.29	-0.247	0.337	-0.73
$ASC_{Shared\ car1}$	-2.980	0.354	-8.43	-3.199	0.352	-9.09
ASC_{Bus1}	-1.140	0.341	-3.35	-1.035	0.331	-3.13
ASC_{Train1}	-2.780	0.682	-4.07	0.839	0.448	1.87
$ASC_{Shared\ taxi1}$	-2.180	0.356	-6.12	-1.705	0.349	-4.89
$ASC_{Walking1}$	-			-		
$ASC_{Train+Bus1}$	-4.930	0.572	-8.62	-3.150	1.214	-2.59
$ASC_{Car\ driver2}$	0.826	0.322	2.56	1.002	0.328	3.05
$ASC_{Shared\ car2}$	-1.670	0.328	-5.09	-1.878	0.334	-5.62
ASC_{Bus2}	-0.659	0.343	-1.92	-0.539	0.336	-1.60
ASC_{Train2}	-1.580	0.47	-3.37	2.014	0.416	4.84
$ASC_{Shared\ taxi2}$	-1.590	0.345	-4.60	-1.110	0.344	-3.23
$ASC_{Walking2}$	1.200	0.168	7.15	1.190	0.174	6.84
$ASC_{Train+Bus2}$	-4.030	0.460	-8.74	-2.237	0.489	-4.57
$\beta_{Cost/Income}$	-0.015	0.003	-5.10	-0.026	0.0067	-3.94
$\beta_{Generalised\ Time\ Car}$	-0.031	0.0075	-4.14	-0.0308	0.0084	-3.80
$\beta_{Generalised\ Time\ Walking}$	-0.113	0.011	-10.35	-0.115	0.011	-10.45
$\beta_{Generalised\ Time\ Public\ transport}$	-0.0075	0.0017	-4.44	-0.0097	0.0017	-5.71
$\beta_{Distance\ travel\ ST1}$	-1.560	0.475	-3.28	-1.952	0.479	-4.08
$\beta_{Distance\ travel\ TTB1}$	2.000	0.645	3.10	-1.568	0.711	-2.21
$\beta_{Distance\ travel\ ST2}$	-1.830	0.376	-4.87	-2.241	0.299	-7.49
$\beta_{Distance\ travel\ TTB2}$	1.660	0.382	4.34	-1.888	0.457	-4.13
$\beta_{\hat{\delta}CI}$	-			0.014	0.007	2.00
$\beta_{\hat{\delta}GT}$	-			0.002	0.0003	6.67
Sample size	2417			2417		
Log-likelihood	-3266.32			-3258.79		

[a] Standard error determined using Bootstrap.

Table 2-1. *Endogenous and corrected mode choice models for Great Valparaíso*

The validity of the instruments was verified using the overidentification tests for the exogeneity of the instruments in DCM proposed by Guevara (2018). In this case, $l(\theta)^{CF}$ is obtained directly from the model in Table 2-1 and $l(\theta)^{CF\_Z}$ was obtained by fixing one of the instruments to zero (for example *IV_GT*) in each case and including the other two instruments (i.e., *IV_C* and *IV_D*) as additional variables within the utility function.

The degree of overidentification for this test is equal to one because the model includes two endogenous variables (Travel Cost and Generalised Time) and three instruments (*IV_GT*, *IV_C* and *IV_D*). It is worth noting that, although the model includes three parameters for Generalised Time, differentiated by mode, the variable is the same. Therefore, it is better to consider it as just one when analysing the degrees of freedom for testing instrument validity. Another alternative, technically also valid, would be to consider it as three different variables, but that would be misleadingly much laxer. Indeed, if we considered Generalised Time as three variables, we should do the same with the respective instrument and, therefore, we would have 1+3 = 4 endogenous variables and 1+6+1 = 8 instruments. Such an approach would lead to four degrees of freedom instead of only one, implying a much laxer critical value of 9.49 instead of 3.84 for the tests shown in (2.16)-(2.18).

The results of the S_{REF} in (2.16), (2.17) and (2.18) show that in all cases $S_{REF} < \chi_1^2$ (3.84); so, we can conclude that all our instruments are indeed exogenous:

$$S_{REF}^{Fixing\ IV\_GT} = -2(-3258.79 + 3258.14) = 1.31 < 3.84 \qquad (2.16)$$

$$S_{REF}^{Fixing\ IV\_C} = -2(-3258.79 + 3258.51) = 0.56 < 3.84 \qquad (2.17)$$

$$S_{REF}^{Fixing\ IV\_D} = -2(-3258.79 + 3257.99) = 1.59 < 3.84 \qquad (2.18)$$

To apply the S_{mREF}, we considered *IV_GT*, *IV_C*, and *IV_D* as additional variables within the DCM, fixed each of the β parameters of the right-side model of Table 2-1 and obtained the log-likelihood $l(\theta)^{CF\_Zall}$ (-3258.27). This give, $S_{mREF} = -2 * (-3258.79 + 3258.27) = 1.05$, less than the critical $\chi_1^2 = 3.84$ value; therefore, we can conclude that all our instruments are valid. To the best of our knowledge, these instruments had not been suggested before to correct for endogeneity in modelling urban mode choice.

The corrected model in Table 2-1 also has parameters for the level-of-service variables with correct signs and statistically significant (at 95% level). An interesting fact is the change of sign in the parameters $\beta_{Distance\ travel\ TTB1}$ and $\beta_{Distance\ travel\ TTB2}$, which suggests that trips over 10 km are actually not preferred to be made by Train and Train+Bus, contrary to the potentially endogenous model. It is also interesting to note that the parameter $\beta_{Cost/Income}$ in the corrected model is 73% higher than the one estimated in the endogenous model.

On the other hand, although the parameters $\beta_{Generalised\ Time\ Car}$ and $\beta_{Generalised\ Time\ Walking}$ are similar in both models, suggesting low bias (less 1%) in their estimation, the parameter $\beta_{Generalised\ Time\ Public\ transport}$ is 30% higher in the corrected model. Thus, the percentage differences between the generalised time parameters for the endogenous and corrected models are smaller in comparison with those of the cost parameter. This result suggests that the cost parameter appears to be more vulnerable to endogeneity than the time parameters and, thereby, it was more poorly estimated in the original model by SECTRA. This finding is in line with that shown by Varela *et al*. (2018). It also suggests that problems such as measurement errors, perception errors and omitted variables, affect more the cost parameter than the time parameters. In practice, then, efforts should focus on improving the way the cost variable is collected and measured in our surveys, to achieve more consistent parameters during model estimation. In particular, the bias in the parameter $\beta_{Generalised\ Time\ Public\ transport}$ can also be due to omitted variables that explain mode choice. If attributes like comfort and reliability, often correlated with travel cost and time, are excluded from the mode choice model, this is a potential source for endogeneity and, as a result, the SVT is overestimated (Tirachini *et al*., 2013). In any case, given the aims of this research we cannot ascertain how much endogeneity is due to some of the sources described previously. It is an interesting research question that is left for future research.

Now, given that the endogenous model is a restricted version of the corrected model, it is possible to apply the likelihood ratio (LR) test (Ortúzar and Willumsen, 2011, page 281) to investigate the presence of endogeneity[4]. The null hypothesis, in this case, is that there is no endogeneity. Then both models are equivalent, rejecting it, implies that the restricted model is erroneous and then endogeneity is present. LR is asymptotically distributed χ_r^2 with r degrees of freedom, where r is the number of linear restrictions required to transform the more general model into the restricted version, which in this case corresponds to the number of residuals incorporated into the corrected model.

In our case, the degrees of freedom of the LR test are $r = 2$ (because the restrictions are that both $\beta_{\hat{\delta}GT}$ and $\beta_{\hat{\delta}CI}$ are zero). So, $LR = -2(-3266.32 + 3258.79) = 15.07$, and this value must be compared with the critical value for two degrees of freedom at the 95% level ($\chi_2^2 = 5.99$). As LR > χ_2^2 the null hypothesis is confidently rejected, and we can conclude that the corrected model is superior.

It must be noted that CF approach tends to yield variances of the estimators that are often larger than those of the *true* model and usually also larger than those of the endogenous model; therefore, its confidence intervals could be wider. Thus, although the correction may be relatively poorer in this regard (at least in some cases), what is crucial is that the estimators will be consistent with the CF correction. Neglecting it may even result in reversing the effect of the attributes due to a change of sign. In the case study analysed, there was no change of sign. Still, the difference in point estimates were as large as 43% in some cases (see Figure 2-1), implying that even if what one cares about is the MSE and not the finite sample bias, the CF results would be preferred. This recommendation is reinforced by the fact that, in strategic transport models the point estimate (i.e., the mean of the estimator distribution) is used for forecasting. For this reason, any bias on the base year values would be exacerbated in future simulations, resulting in poor model forecasting performance.

2.3.4. *Quantification of effects due to endogeneity*

In this subsection we quantify the impacts of endogeneity in the model. The measures used for this are SVT and aggregate Elasticities. These were calculated for the endogenous and corrected model and later compared. To estimate each of these measures, we divided the dataset into two samples: 80% for estimation and 20% for validation (holdout sample). Additionally, the process was repeated 100 times (i.e., 100 repetitions), with the aim of guaranteeing randomness in the estimates.

It should be noted that the estimation of the standard errors when applying the two-stage CF method comes with a caveat. Given that the proposed estimator is estimated in two stages,

[4] Following Rivers and Vuong (1988), note that when using a two-step procedure, the test for the presence of endogeneity does not need correcting the standard errors with bootstrap. This holds because the test is evaluated under the null hypothesis that there is no endogeneity. Therefore, the population coefficient of the residuals is zero. This logic holds for Wald, Lagrange Multiplier and LR tests, when used to evaluate the presence of endogeneity, which is what we use in this section (see, for example, the discussion in Guevara, 2010, Ch. 2).

variances cannot be calculated directly from the Fisher-information-matrix. Therefore, to make the inference, the variance-covariance matrix must be determined using nonparametric methods such as bootstrapping (Petrin and Train 2002) or the approach proposed by Karaca-Mandic and Train (2003), or by writing, instead, the full likelihood of both stages together (Train 2003).

In this application, we used the bootstrap approach to calculate the standard errors and the confidence intervals for the estimators reported in Table 2-1 and for the SVT and elasticities reported in Tables 2-2, 2-3 and 2-4. Confidence intervals for both SVT and elasticities were estimated using the percentile bootstrap method (Davison and Hinkley 1997). This approach considers using the percentiles of the bootstrap distribution directly (in our case 2.5% and 97.5%), to represent a confidence interval at the 5% significance level.

Model	**Private**	**Public**
Endogenous	2.05 (1.02 - 4.19)	0.49 (0.31 - 0.98)
Corrected	1.17 (0.18 - 2.16)	0.37 (0.17 - 0.63)

[b] Confidence interval (in parenthesis)

Table 2-2. *Mean and confidence intervals for the SVT*[b]

Model \ Mode	**Car driver1**	**Shared car1**	**Bus1**	**Train1**	**Shared taxi1**	**Walking1**	**Train+Bus1**
Endogenous	-0.347 (-0.533 – -0.211)	-0.413 (-0.632 – -0.257)	-0.504 (-0.797 – -0.358)	-0.751 (-1.187 – -0.532)	-0.523 (-0.858 – -0.384)	-1.96 (-2.258 – -1.692)	-0.807 (-1.266 – -0.559)
Corrected	-0.346 (-0.422 – -0.063)	-0.436 (-0.517 – -0.088)	-0.657 (-0.781 – -0.339)	-0.974 (-1.164 – -0.502)	-0.665 (-0.846 – -0.368)	-1.998 (-2.248 – -1.681)	-1.074 (-1.26 – -0.55)
Model \ Mode	**Car driver2**	**Shared car2**	**Bus2**	**Train2**	**Shared taxi2**	**Walking2**	**Train+Bus2**
Endogenous	-0.268 (-0.408 – -0.162)	-0.468 (-0.692 – -0.295)	-0.451 (-0.72 – -0.317)	-0.555 (-0.934 – -0.414)	-0.48 (-0.836 – -0.371)	-1.336 (-1.502 – -1.134)	-0.771 (-1.218 – -0.537)
Corrected	-0.269 (-0.325 – -0.053)	-0.494 (-0.568 – -0.101)	-0.588 (-0.708 – -0.302)	-0.722 (-0.853 – -0.368)	-0.604 (-0.823 – -0.354)	-1.365 (-1.492 – -1.128)	-1.029 (-1.212 – -0.528)

[c] Confidence interval (in parenthesis)

Table 2-3. *Mean and confidence intervals for the Generalised Time elasticities*[c]

Model \ Mode	**Car driver1**	**Bus1**	**Train1**	**Shared taxi1**	**Train+Bus1**
Endogenous	-0.318 (-0.422 – -0.207)	-0.082 (-0.114 – -0.051)	-0.089 (-0.118 – -0.052)	-0.145 (-0.213 – -0.097)	-0.111 (-0.153 – -0.066)
Corrected	-0.54 (-0.821 – -0.304)	-0.139 (-0.216 – -0.076)	-0.15 (-0.232 – -0.084)	-0.241 (-0.411 – -0.143)	-0.192 (-0.293 – -0.102)
Model \ Mode	**Car driver2**	**Bus2**	**Train2**	**Shared taxi2**	**Train+Bus2**
Endogenous	-0.196 (-0.261 – -0.125)	-0.066 (-0.091 – -0.042)	-0.071 (-0.1 – -0.045)	-0.133 (-0.202 – -0.092)	-0.112 (-0.159 – -0.068)
Corrected	-0.333 (-0.516 – -0.189)	-0.112 (-0.174 – -0.063)	-0.12 (-0.19 – -0.066)	-0.22 (-0.388 – -0.136)	-0.197 (-0.304 – -0.105)

[d] Confidence interval (in parenthesis)

Table 2-4. *Mean and confidence intervals for the Cost/Income elasticities*[d]

Given that the cost variable is really Cost/Income, then SVT is expressed as [% Income/min]. It was possible to estimate it separately for the private and public transport modes, given that the Generalised Time parameters were specific for these modes (see Figure 2-1). As can be seen, the SVT of the original model was overestimated in comparison with that obtained for the corrected model. For the private modes, the SVT suffers an overestimation of up to 43%, while in the case of the public transport modes this reaches 26%. These findings are in line with those shown by Varela *et al*. (2018), who used a case study for Stockholm commuters to assess the magnitude of the measurement errors in travel time and travel cost using latent variables. These differences are important, because measures such as SVT are critical in the social evaluation of transport projects. Given that the bias of the cost parameter is higher than the bias of the time parameters, it makes sense that the SVT estimates are overestimated. If the SVT is biased, the social evaluation of the project will likely be biased too.

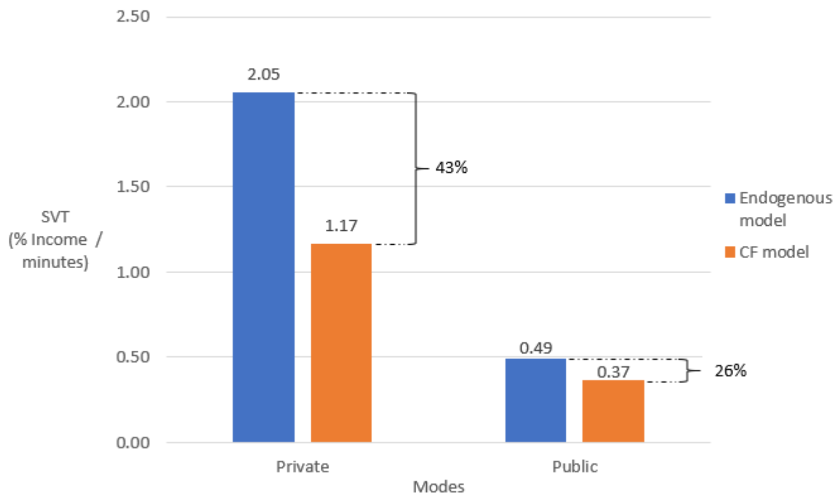

Figure 2-1. *SVT for private and public modes*

Elasticities are frequently used in transport project evaluation (Ortúzar and Willumsen 2011). In the case of the elasticities of the Generalised Times (Figure 2-2) and Cost/Income (Figure 2-3) for the original model, we can see that these are underestimated in comparison with those obtained using the corrected model. This finding is also consistent with the results of Varela *et al*. (2018) and Varotto *et al*. (2017), who observed increases of up to 65% in the time elasticity value and of up to 50% of the price elasticity, when assessing the magnitude of the measurement errors in these variables (using latent variables), in a large-scale travel demand model.

In our case, the Generalised Time elasticities are underestimated up to 33%, while the Cost/Income elasticities are underestimated up to 75%. The mode with the highest generalised time elasticity is Walking (in both periods). On the other hand, the smallest generalised time elasticity is registered for Car driver (in both periods), but the generalised time elasticities for both private modes (Car driver and Shared car) show no differences between the endogenous and the corrected model.

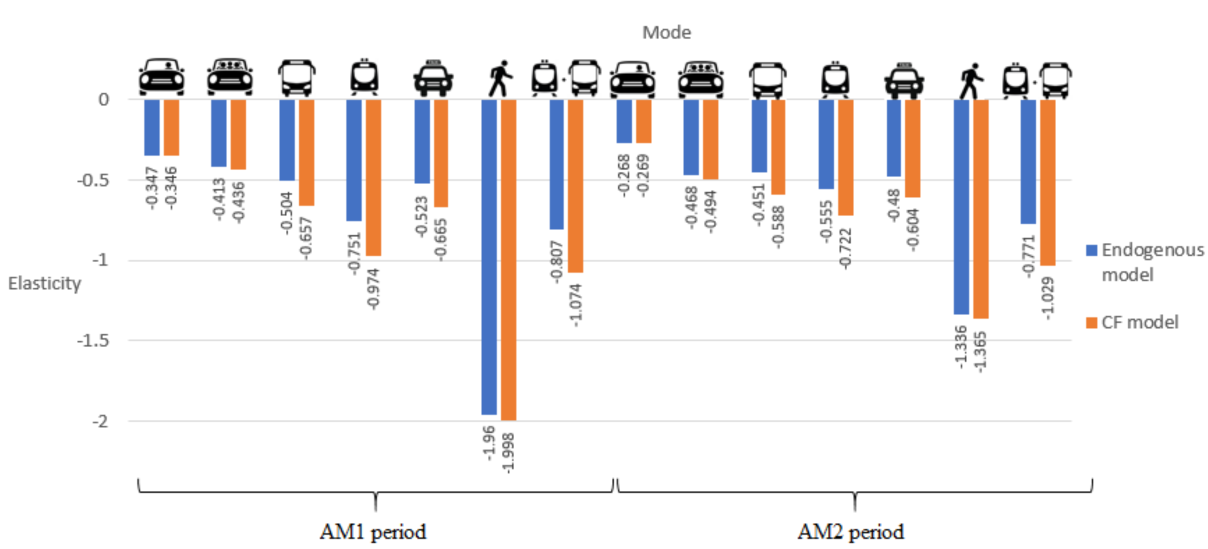

Figure 2-2. *Generalised Time elasticities*

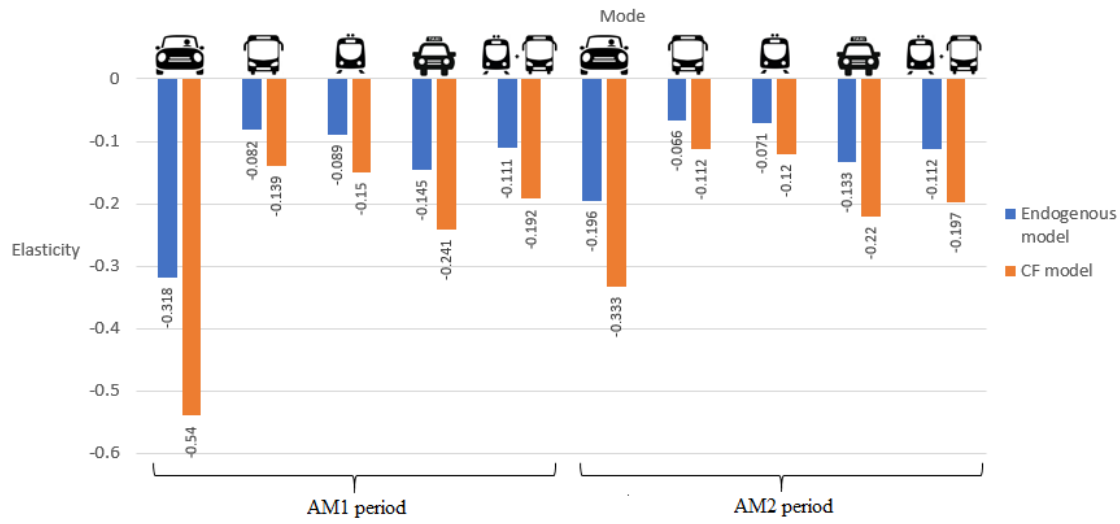

Figure 2-3. *Cost/Income elasticities*

Cost/Income elasticities for the AM1 period (-0.54) are higher in the Car driver mode than in the public transport modes (where they vary between -0.139 to -0.241). These results are also consistent with findings from other studies (Varela *et al*. 2018). Given that the parameter $\beta_{Cost/Income}$ and $\beta_{Generalised\ Time\ Public\ transport}$ were underestimated in the endogenous model, the underestimation of elasticities was expected. Note that the Generalised time elasticities calculated for both models and for the modes Car driver, Shared car and Walking, are also similar because both parameters $\beta_{Generalised\ Time\ Car}$ and $\beta_{Generalised\ Time\ Walking}$ have a rather low bias.

2.4. Conclusions and Future Research Directions

Endogeneity is an anomaly that also arises in urban mode choice models at the strategic level. It affects the consistency of the model parameters estimated, especially those related to the travel cost and travel time variables. As these are key explanatory variables in strategic mode choice models, not correcting the endogenous models may lead to faulty decision-making.

The research described in this chapter provides a framework that uses the CF method to correct for the endogeneity of mode choice models at the strategic level using appropriate instrumental variables. The CF method can be considered an adequate methodology in this case. The instruments used were: (*i*) The average travel time of other origin-destination pairs that have a similar length than the origin and destination of the considered trip; (*ii*) the average travel cost of other origin-destination pairs that have a similar length than the origin and destination of the considered trip, and (*iii*) the network trip distance between the origin and the destination for each mode. Defining these instruments is a relevant finding, and they can be considered valid. Government planning agencies (central or local) should begin to consider the CF approach and the instruments used in this research as a guide to correct mode choice models that may present endogeneity.

The confidence in strategic urban mode choice models based on level-of-service variables, such as travel cost and travel time, must be questioned. Our results show that the cost parameters could be more poorly estimated than the time parameters. This may be due to the fact that urban mode choice models at the strategic level may be affected by three sources of endogeneity: measurement errors, omitted variables and simultaneous estimation. We recommended: (*i*) to use instruments within the framework shown in this chapter to improve the estimations, and (*ii*) to focus the efforts in improving the way the cost variable is collected and measured in surveys, to achieve more consistent parameters during model estimation.

We quantified the effects of endogeneity in strategic urban mode choice models. We found that the SVT was overestimated by 43% and 26% for private and public modes, respectively in our case study. This fact may have a strong influence in the social evaluation of transport projects where the SVT is critical. We also showed the impact on model elasticities, finding that these were underestimated. In particular, the Generalised Time elasticities showed underestimations of up to 33%, while the Cost/Income elasticities reached underestimations of up to 75%.

Three areas for further research can be identified. First, we believe it is important to study how correcting for endogeneity would work in forecasting when the variables that change are endogenous, such as travel times and cost in a strategic transport planning model. We also recommend examining in greater depth how the social evaluation of transport projects may be affected by endogeneity, especially given our findings regarding the changes in SVT. Finally, an exciting topic for further research is the identification of weak and strong instruments for correcting endogeneity, because this has been solved for linear models (Stock and Yogo, 2005) but not yet fully extended for DCM.

3. Forecasting with Strategic Transport Models Corrected for Endogeneity

3.1. Introduction

In this chapter we address the important question of how to make forecasts with such models in the case of strategic supply-demand equilibration settings. To address this limitation, we suggest a new approach, the *Control Function Updated* (*CFU*) method, which allows correcting for endogeneity when models are used to forecast demand after changes in level-of-service variables determined at equilibrium for future scenarios (e.g., 10 to 40 years ahead from the calibration year). Our approach takes as a point of reference the *CF* approach of Guevara and Ben-Akiva (2012). This approach is used first to correct for endogeneity in the base year. Then, the correction for future scenarios (where new supply-demand equilibria are reached) consists in updating the *CF* approach used in the first stage, given that the instruments might change in future scenarios (i.e., new equilibria are reached for demand increases and higher flows in the network) and, therefore, affect the control function, which needs to be updated. To test our proposed approach, we used simulated data. We considered three typical sources of endogeneity that may occur in strategic transport models: (*i*) measurement error, (*ii*) omitted variables and (*iii*) the simultaneous estimation of key variables in a supply-demand equilibration mechanism. We also considered six transport modes, in an attempt to emulate the application of a strategic transport model in an urban case. Future scenarios were assessed considering exogenous changes in the explanatory variables.

In this setting, we compared three different approaches: (*i*) do nothing (i.e., no endogeneity correction), (*ii*) the *CF* approach of Guevara and Ben-Akiva (2012) and (*iii*) our proposed *CFU* procedure. The forecasts were evaluated in terms of the recovery of the *true* (simulated) travel time, the logarithm of the likelihood expected value - $E(l(\theta))$ and the Akaike Information Criteria's expected value - $E(AIC)$ (Akaike 1974) for the future scenarios 10 to 40 years ahead.

The chapter is structured as follows. In Section 3.2, we describe the *CF* method's fundamentals and our approach (*CFU*) to correct for endogeneity in DCM. In Section 3.3, we describe the Monte Carlo experiment designed to test our methodological proposal (*CFU*) to correct for endogeneity, when forecasting the demand after changes in level-of-service variables determined at equilibrium. The results are shown and analysed in Section 3.4. Finally, in Section 3.5, we conclude summarising the main findings and suggesting future lines of research.

3.2. Addressing Endogeneity Using the CF Approach and the CFU Method

3.2.1. *CF approach*

The *CF* approach's idea is attributed to Heckman (1978), who first developed it to correct endogeneity in a simultaneous equation problem. The *CF* classical version can be applied following a two-stage procedure or simultaneously (Train 2009). To explain the method's fundamentals, consider a DCM with a utility function represented by (3.1):

$$U_{in} = ASC_i + \beta_t t_{in} + \beta_c c_{in} + \beta_q q_{in} + \varepsilon_{in} \qquad (3.1)$$

where U_{in} represents the utility perceived by individual *n* for alternative A_i belonging to the individual's choice set *A(n)*. ASC_i is an alternative specific constant for alternative A_i. β_t and β_c are parameters associated with the explanatory variables t_{in} and c_{in} (representing, for example, travel time and cost). β_q is a parameter associated with the qualitative attribute q_{in} (e.g., safety, comfort, or reliability). Finally, ε_{in} is an exogenous error term for individual *n* and alternative A_i.

Given the above, let us assume that for illustrative purposes of this explanation the qualitative attribute q_{in} is unknown to the modeller; the modeller's specification will be as in (3.2), where the new error term $\tilde{\varepsilon}_{in}$ contains both ε_{in} and q_{in}:

$$U_{in} = ASC_i + \beta_t t_{in} + \beta_c c_{in} + \tilde{\varepsilon}_{in} \qquad (3.2)$$

For explanatory purposes, we will also consider that from (3.2) the explanatory variables c_{in} and t_{in} are endogenous due to three possible sources of endogeneity: measurement errors, omitted variables, simultaneous determination. Guerrero *et al*. (2020) showed that these variables are usually endogenous in strategic mode choice models. We assume that the endogeneity for c_{in} is due to measurement error and omitted variables, whereas the endogeneity for t_{in} is due to its simultaneous estimation process.

For the explanatory variable c_{in}, we consider that it is correlated with q_{in} as shown in (3.3). where, θ_c is a constant, z_{in} is an exogenous attribute, which then works as instrument, since it partially explains c_{in}. θ_z is parameter to be estimated. $\tilde{\varphi}_{in}$ is an error term containing both φ_{in} and q_{in}. The correlation between c_{in} and q_{in} in (3.3) yields that c_{in} and the error term $\tilde{\varepsilon}_{in}$ are also correlated in (3.2); therefore, by definition, c_{in} is endogenous. As can be seen, in this case the endogeneity arises due to the omission of the attribute q_{in} in (3.2):

$$c_{in} = \theta_c + \theta_z z_{in} + \theta_q q_{in} + \varphi_{in} = \theta_c + \theta_z z_{in} + \tilde{\varphi}_{in} \qquad (3.3)$$

On the other hand, to account for the measurement error, let us assume that the modeller observes $\tilde{c}_{in}$ instead of c_{in} as shown in (3.4), where η_{in} represents a measurement error and

$\tilde{c}_{in}$ is the explanatory variable but affected by η_{in}. Therefore, c_{in} can be re-written as $c_{in} = \tilde{c}_{in} - \eta_{in}$ and the new functional form in (3.2) changes as follows in (3.5):

$$\tilde{c}_{in} = c_{in} + \eta_{in} \tag{3.4}$$

$$U_{in} = ASC_i + \beta_t t_{in} + \beta_c(\tilde{c}_{in} - \eta_{in}) + \tilde{\varepsilon}_{in} = ASC_i + \beta_t t_{in} + \beta_c \tilde{c}_{in} - \underbrace{\beta_c \eta_{in} + \tilde{\varepsilon}_{in}}_{\tilde{\tilde{\varepsilon}}_{in}} \tag{3.5}$$

where the new error term ($\tilde{\tilde{\varepsilon}}_{in}$) contains both $\beta_c \eta_{in}$ and $\tilde{\varepsilon}_{in}$. As can be seen, the endogeneity arises in (3.5) given that $\tilde{c}_{in}$ is correlated with $\tilde{\tilde{\varepsilon}}_{in}$ through η_{in} and q_{in}. Namely, the expected value of $\tilde{\tilde{\varepsilon}}_{in}$ conditional on $\tilde{c}_{in}$ is not zero ($E(\tilde{\tilde{\varepsilon}}_{in}|\tilde{c}_{in}) \neq 0$), therefore $\tilde{c}_{in}$ is endogenous.

To represent endogeneity in the explanatory variable t_{in}, let us consider the supply-demand equilibrium mechanism that occurs in the classical four-stage transport model. In this case, new equilibria are reached due to changes in demand, which in turn yield changes on network flows (supply). Namely, the error term $\tilde{\tilde{\varepsilon}}_{in}$ affects the probability (P_{in}) in (3.6), which triggers a change in $\tilde{t}_{in}$. For explanatory purpose, $\tilde{t}_{in}$is expressed through of a BPR function (Bureau of Public Roads 1964) as shown in (3.7), where fft_{in} represents the free-flow time, and μ and ρ are parameters to be estimated. The values reached as a result of this equilibrium will be known as $\tilde{t}_{in}$, and they are considered endogenous, because the expected value of $\tilde{\tilde{\varepsilon}}_{in}$ conditional on $\tilde{t}_{in}$, $E(\tilde{\tilde{\varepsilon}}_{in}|\tilde{t}_{in}) \neq 0$.

$$P_{in} = \frac{e^{ASC_i + \beta_t \tilde{t}_{in} + \beta_c \tilde{c}_{in} + \tilde{\tilde{\varepsilon}}_{in}}}{\sum_{j \epsilon A} e^{ASC_j + \beta_t \tilde{t}_{jn} + \beta_c \tilde{c}_{jn} + \tilde{\tilde{\varepsilon}}_{jn}}} \tag{3.6}$$

$$\tilde{t}_{in} = fft_{in}[1 + \mu(P_{in})^{\rho}] \tag{3.7}$$

As mentioned above, an essential requirement for using the *CF* method to correct for endogeneity is the availability of proper instruments (Hausman, 1978). An instrument can be considered valid when fulfilling two requirements: (*i*) it is correlated with the endogenous variable, and (*ii*) it is independent of the model's error term. The former is known as *relevance condition* and the second as *exogeneity condition*. There are some tests in the literature to determine whether both conditions are fulfilled for the case of DCM. The relevance condition can be checked using the preliminary results based on the F test of the first stage regression of the *CF* method, as proposed by Guevara and Navarro (2015). To check the exogeneity condition, Guevara (2018) recently proposed the *Refutability* test and its variation, the *Modified Refutability* test. These can be applied when there is overidentification of the model, which holds when there are more instruments than endogenous variables. If the instruments are relevant and exogenous, the *CF* method can be used to obtain consistent estimators of the model parameters, up to a scale. Nevertheless, it is always a challenge (and even a controversial one, see Bresnahan, 1997) to find proper instruments fulfilling both conditions. To apply the *CF* method, at least one instrument is needed for each endogenous variable considered.

In practice, following the two-stage approach of the *CF* method (Train, 2009; Wooldridge, 2010), the first stage below consists in obtaining the residuals ($\hat{\delta}_{in}$) from an ordinary least square (OLS) regression of $\tilde{t}_{in}$ and $\tilde{c}_{in}$ on the exogenous variables in the DCM and the instruments. One control function must be estimated for each endogenous variable. Therefore, the first stage of the *CF* approach for $\tilde{t}_{in}$ and $\tilde{c}_{in}$ is as follows:

$$\tilde{t}_{in} = \alpha_t + \gamma_z^t z_{in} + \gamma_{fft}^t fft_{in} + \delta_{in}^t \quad \overset{OLS}{\longrightarrow} \quad \hat{\delta}_{in}^t = \tilde{t}_{in} - \hat{\tilde{t}}_{in} \qquad (3.8)$$

$$\tilde{c}_{in} = \alpha_c + \gamma_z^c z_{in} + \gamma_{fft}^c fft_{in} + \delta_{in}^c \quad \overset{OLS}{\longrightarrow} \quad \hat{\delta}_{in}^c = \tilde{c}_{in} - \hat{\tilde{c}}_{in} \qquad (3.9)$$

where, α_c and α_t are the intercepts of the regression model, γ are parameters to be estimated, and δ are the regression model's error terms. On the other hand, the instruments (i.e., exogeneous variables) considered are z_{in} and fft_{in} given that they are correlated with the endogenous variable, but they do not influence the individuals' choice.

In the second stage, the DCM is estimated considering $\hat{\delta}_{in}^t$ and $\hat{\delta}_{in}^c$ (coming from the first stage) as explanatory variables within the utility function, as shown in (3.10):

$$U_{in} = ASC_i + \beta_t \tilde{t}_{in} + \beta_c \tilde{c}_{in} + \beta_{\hat{\delta}^t} \hat{\delta}_{in}^t + \beta_{\hat{\delta}^c} \hat{\delta}_{in}^c + \tilde{\tilde{\varepsilon}}_{in} \qquad (3.10)$$

The intuition behind this method is that the residuals $\hat{\delta}_{in}^t$ and $\hat{\delta}_{in}^c$ capture all of $\tilde{t}_{in}$ and $\tilde{c}_{in}$ that is correlated with the error term $\tilde{\varepsilon}_{in}$, effectively controlling for this problem in (3.10). However, this approach implies a different error term of the model ($\tilde{\tilde{\varepsilon}}_{in}$ instead of $\tilde{\varepsilon}_{in}$), resulting in a change of scale in the estimated parameters of the DCM (Yatchew and Griliches 1985; Cramer 2007; Guevara and Ben-Akiva, 2012). Thereby, it would be wrong to check the recovery of the parameter for $\tilde{t}_{in}$ and $\tilde{c}_{in}$ in (3.10); rather, its ratio (i.e. β_t/β_c) needs to be checked.

Now, while the two-stage procedure guarantees consistent estimators up to a scale, their standard errors cannot be calculated, in general, directly from the Fisher information matrix due to the estimated regressor problem (Guevara, 2015). Therefore, they must be corrected to carry out statistical inference analysis. The only case in which this correction is not needed is for testing that $\beta_{\hat{\delta}^t}$ and $\beta_{\hat{\delta}^c}$ in (3.10) are equal to zero, because under such a null assumption there is no endogeneity. Therefore, this can be used as a direct test to discard the presence of endogeneity (Rivers and Vuong, 1988). Other alternatives to correct the standard errors are the Bootstrap method (Petrin and Train, 2003) or the formula proposed by Hardin (2002), by Karaca-Mandic and Train (2003), or the one proposed by Terza (2016). If the modeller wants to avoid the standard error correction, the maximum likelihood simultaneous *CF* approach should be applied (Train, 2009).

Furthermore, the two-stage estimation approach results in an efficiency loss, which can be avoided when the error terms $\tilde{\varepsilon}_{in}$ and $\tilde{\varphi}_{in}$ are homoscedastic and non-autocorrelated (Rivers and Vuong, 1988). For other cases, efficiency may only be regained by following a maximum likelihood approach, as suggested by Train (2009).

Note finally, that the mathematical derivation of the *CF* approach considers that the error terms $\tilde{\tilde{\varepsilon}}_{in}$ and $\tilde{\varphi}_{in}$ distribute bivariate Normal. Therefore, the specification in (3.1) leads to a Probit model. However, Villa-Boas (2007) showed that, under some considerations, approximating it to a Gumbel distribution (also called Extreme Value Type I) did not involve any issue.

Guevara and Ben-Akiva (2012) considered the question of how to forecast and simulate in practice using the *CF* method under exogenous shifts. Formally, the estimation of the probability P_{in}^{1} given Gumbel errors corresponds to the following multinomial logit expression:

$$P_{in}^{1} = \frac{e^{\widehat{ASC}_i + \hat{\beta}_t \tilde{t}_{in}^{1} + \hat{\beta}_c \tilde{c}_{in}^{1} + \hat{\beta}_{\hat{\delta}^t} \hat{\delta}_{in}^{t} + \hat{\beta}_{\hat{\delta}^c} \hat{\delta}_{in}^{c}}}{\sum_{j \in A} e^{\widehat{ASC}_j + \hat{\beta}_t \tilde{t}_{jn}^{1} + \hat{\beta}_c \tilde{c}_{jn}^{1} + \hat{\beta}_{\hat{\delta}^t} \hat{\delta}_{jn}^{t} + \hat{\beta}_{\hat{\delta}^c} \hat{\delta}_{jn}^{c}}} \tag{3.11}$$

where the parameters $\widehat{ASC}_i$ and $\hat{\beta}$ in (3.11) are estimators obtained following the two-stage *CF* approach shown previously. Here, the superscript 1 is used to highlight that the model attributes vary in the forecasting phase; however, $\hat{\delta}_{in}$ is fixed. This last comment is crucial because unlike Guevara and Ben-Akiva (2012), we hypothesize that in future scenarios (where new supply-demand equilibria are reached) the *CF* approach's first stage must be updated. Namely, the instruments and level-of-service attributes might endogenously change in future scenarios. If this is not corrected, it may cause biases as we will show later.

On the other hand, if the *CF* approach's simultaneous (one-stage) procedure is preferred, forecasting would follow (3.12):

$$P_{in}^{1} = \frac{e^{\widehat{ASC}_i + \hat{\beta}_t \tilde{t}_{in}^{1} + \hat{\beta}_c \tilde{c}_{in}^{1} + \hat{\beta}_{\hat{\delta}^t} (\tilde{t}_{in} - \hat{\alpha}_t - \hat{\gamma}_z^t z_{in}^0 - \hat{\gamma}_{fft}^t fft_{in}^0) + \hat{\beta}_{\hat{\delta}^c} (\tilde{c}_{in} - \hat{\alpha}_c - \hat{\gamma}_z^c z_{in}^0 - \hat{\gamma}_{fft}^c fft_{in}^0)}}{\sum_{j \in A} e^{\widehat{ASC}_j + \hat{\beta}_t \tilde{t}_{jn}^{1} + \hat{\beta}_c \tilde{c}_{jn}^{1} + \hat{\beta}_{\hat{\delta}^t} (\tilde{t}_{jn} - \hat{\alpha}_t - \hat{\gamma}_z^t z_{jn}^0 - \hat{\gamma}_{fft}^t fft_{jn}^0) + \hat{\beta}_{\hat{\delta}^c} (\tilde{c}_{jn} - \hat{\alpha}_c - \hat{\gamma}_z^c z_{jn}^0 - \hat{\gamma}_{fft}^c fft_{jn}^0)}} \tag{3.12}$$

where, superscript 0 indicates data coming from the sample used for estimation and superscript 1 indicates attributes that vary in the forecasting phase (Guevara and Ben-Akiva 2012).

3.2.2. *CFU method*

The approach of Guevara and Ben-Akiva (2012) has a potential limitation in forecasting because it assumes that exogenous shifts do not affect predictions in future scenarios. In strategic transport models, the simulation of future scenarios requires achieving new supply/demand equilibria, which must be considered endogenous as they result from an equilibrium process. This might involve a change, not necessarily linear, of the relationship between the residuals and the endogenous variables, invalidating the use of (3.11) or (3.12).

Namely, the endogeneity continues arising in the future years affecting the calculation of the probabilities; however, there is no guarantee that the instruments used for the calibration year correct it, given that the correlation between the instruments (z_{in} and fft_{in}) and the endogenous variable ($\tilde{c}_{in}$ and $\tilde{t}_{in}$) changes in (3.8) and (3.9). This fact is ignored by Guevara and Ben-Akiva (2012), and all related literature since they are only concern with the analysis of exogenous shifts. In practice, these situations can occur - for example - when the free-flow times change due to improvements in the infrastructure or when the fuel cost changes due to governmental policies. Note that for both cases, the shifts are exogenous, but it triggers endogenous changes in the forecasting of the future scenarios. We propose to address this limitation by updating the residuals for applying the *CF* method for futures scenarios. We termed this proposed variation of the *CF* method as Control Function Updated (*CFU*), which will be explained as follow.

For explanatory purposes, we continue using the superscript 0 to indicate data coming from the sample used for estimation and the superscript 1 to indicate attributes that vary in the forecasting phase. The proposed *CFU* method for forecasting consists in considering that the residuals from the first stage of the *CF* approach for the future scenarios must be obtained using the value of the instruments for the year of forecasting (z_{in}^1 and fft_{in}^1) instead of z_{in}^0 and fft_{in}^0 in (3.8) and (3.9). Therefore, these new residuals are identified as $\hat{\delta}_{in}^{1t}$ and $\hat{\delta}_{in}^{1c}$. They are obtained applying an OLS regression of $\tilde{c}_{in}^1$ and $\tilde{t}_{in}^1$ on the exogenous variables in the DCM and the instruments. In this way, the CFU approach's first stage is represented in (3.13) and (3.14):

$$\tilde{t}_{in}^1 = \alpha_t + \gamma_z^t z_{in}^1 + \gamma_{fft}^t fft_{in}^1 + \delta_{in}^{1t} \quad \xrightarrow{OLS} \quad \hat{\delta}_{in}^{1t} = \tilde{t}_{in}^1 - t_{in}^1 \qquad (3.13)$$

$$\tilde{c}_{in}^1 = \alpha_c + \gamma_z^c z_{in}^1 + \gamma_{fft}^c fft_{in}^1 + \delta_{in}^{1c} \quad \xrightarrow{OLS} \quad \hat{\delta}_{in}^{1c} = \tilde{c}_{in}^1 - c_{in}^1 \qquad (3.14)$$

where, α is the intercept of the regression model, γ_z and γ_{fft} are parameters to be estimated for the exogenous attributes z_{in}^1 and fft_{in}^1, respectively. Finally, δ_{in}^1 is the error term of the regression model.

For the second stage of the CFU approach, $\hat{\delta}_{in}^{1c}$ and $\hat{\delta}_{in}^{1t}$ are added as explanatory variables within the utility function. However, the choices are unknown; therefore, we use P_{in}^1 coming from (3.11) or (3.12) as a proxy of the choice. Furthermore, it is known that forecasting in transport planning and social evaluation projects requires the parameters estimated for the calibration year (i.e., $\widehat{ASC}_i$, $\hat{\beta}_t$ and $\hat{\beta}_c$). The attribute vectors ($\tilde{t}_{in}^1$, $\tilde{c}_{in}^1$, $\hat{\delta}_{in}^{1c}$ and $\hat{\delta}_{in}^{1t}$) are known for the modeller, therefore the only unknown elements are $\hat{\beta}_{\hat{\delta}^{1c}}$ and $\hat{\beta}_{\hat{\delta}^{1t}}$. These can be re-estimated using the linear regression in (3.15), which can be seen as an application of the Berkson-Theil transformation procedure (Ortúzar and Willumsen 2011).

$$Ln\frac{P_{in}^1}{P_{jn}^1} = \left(\widehat{ASC}_i - \widehat{ASC}_j\right) + \hat{\beta}_t\left(\tilde{t}_{in}^1 - \tilde{t}_{jn}^1\right) + \hat{\beta}_c\left(\tilde{c}_{in}^1 - \tilde{c}_{jn}^1\right) + \beta_{\hat{\delta}^{1c}}\left(\hat{\delta}_{in}^{1c} - \hat{\delta}_{jn}^{1c}\right) + \beta_{\hat{\delta}^{1t}}\left(\hat{\delta}_{in}^{1t} - \hat{\delta}_{jn}^{1t}\right) + \left(\tilde{\tilde{\tilde{\varepsilon}}}_{in} - \tilde{\tilde{\tilde{\varepsilon}}}_{jn}\right) \qquad (3.15)$$

where the left-hand side of (3.15) is estimated based on Guevara and Ben-Akiva (2012) approach and it acts as the dependent variable; whereas $\left(\hat{\delta}_{in}^{1c} - \hat{\delta}_{jn}^{1c}\right)$ and $\left(\hat{\delta}_{in}^{1t} - \hat{\delta}_{jn}^{1t}\right)$ act as the independent variables. Then, the sum of $\hat{\beta}_t\left(\tilde{t}_{in}^1 - \tilde{t}_{jn}^1\right)$, $\hat{\beta}_c\left(\tilde{c}_{in}^1 - \tilde{c}_{jn}^1\right)$ and $\left(\widehat{ASC}_i - \widehat{ASC}_j\right)$ is the intercept.

The calculation of the probability $P_{in}^{1\_CFU}$ after updating the residuals is given by (3.16) for the two-stage *CF* approach. Note that the parameters used are $\widehat{ASC}_i$, $\hat{\beta}_t$ and $\hat{\beta}_c$ and come from the calibration year model, however $\hat{\beta}_{\hat{\delta}^{1c}}$ and $\hat{\beta}_{\hat{\delta}^{1t}}$ are used instead $\beta_{\hat{\delta}^c}$ and $\beta_{\hat{\delta}^t}$. On the other hand, if the one-stage estimation is used, then it must follow (3.17):

$$P_{in}^{1\_CFU} = \frac{e^{\widehat{ASC}_i + \hat{\beta}_t \tilde{t}_{in}^1 + \hat{\beta}_c \tilde{c}_{in}^1 + \hat{\beta}_{\hat{\delta}^{1t}} \hat{\delta}_{in}^{1t} + \hat{\beta}_{\hat{\delta}^{1c}} \hat{\delta}_{in}^{1c}}}{\sum_{j \in A} e^{\widehat{ASC}_j + \hat{\beta}_t t_{jn}^1 + \hat{\beta}_c c_{jn}^1 + \hat{\beta}_{\hat{\delta}^{1t}} \hat{\delta}_{jn}^{1t} + \hat{\beta}_{\hat{\delta}^{c1}} \hat{\delta}_{jn}^{1c}}} \qquad (3.16)$$

$$P_{in}^{1\_CFU} = \frac{e^{\widehat{ASC}_i + \hat{\beta}_t \tilde{t}_{in}^1 + \hat{\beta}_c \tilde{c}_{in}^1 + \hat{\beta}_{\hat{\delta}^{1t}}(\tilde{t}_{in}^1 - \hat{\alpha}_t - \hat{\gamma}_z^t z_{in}^1 - \hat{\gamma}_{fft}^t fft_{in}^1) + \hat{\beta}_{\hat{\delta}^{1c}}(\tilde{c}_{in}^1 - \hat{\alpha}_c - \hat{\gamma}_z^c z_{in}^1 - \hat{\gamma}_{fft}^c fft_{in}^1)}}{\sum_{j \in A} e^{\widehat{ASC}_j + \hat{\beta}_t \tilde{t}_{jn}^1 + \hat{\beta}_c \tilde{c}_{jn}^1 + \hat{\beta}_{\hat{\delta}^{1t}}(\tilde{t}_{jn}^1 - \hat{\alpha}_t - \hat{\gamma}_z^t z_{jn}^0 - \hat{\gamma}_{fft}^t fft_{jn}^0) + \hat{\beta}_{\hat{\delta}^{1c}}(\tilde{c}_{jn}^1 - \hat{\alpha}_c - \hat{\gamma}_z^c z_{jn}^0 - \hat{\gamma}_{fft}^c fft_{jn}^0)}} \qquad (3.17)$$

3.3. A Monte Carlo Framework to Represent Simulated Equilibrium and Forecasting with Strategic Transport Models Corrected for Endogeneity

To investigate the problem at hand, we designed a Monte Carlo experiment. The simulation process considered a sample size of 5000 individuals and was repeated 100 times, to guarantee randomness in the estimates. All experiment runs were generated using the open-source software R (R Development Core Team 2008). We considered six transport modes: Bus, Car driver, Shared car, Walking, Train and Shared taxi, emulating the typical alternatives considered in the application of a strategic 4-stage urban transport model; we also assumed that the availability of modes could vary among individuals (4 to 6 alternatives available). On the other hand, the network structure (i.e., square grid, monocentric, radial, or other) was not considered. Besides, we did not consider a particular number of OD pairs either. This could be interpreted as each observation being an OD pair isolated from the others and with different free-flow times.

For the choice's simulation process, we specified six utility functions shown in (3.18)-(3.23), one for each simulated transport mode:

$$U_{Bus,n} = \beta_c * c_{Bus,n} + \beta_t^{PT} * t_{Bus,n} + \beta_q * q_{Bus,n} + \varepsilon_{Bus,n} \qquad (3.18)$$

$$U_{Car\ driver,n} = \beta_c * c_{Car\ driver,n} + \beta_t^{CM} * t_{Car\ driver,n} + \beta_q * q_{Car\ driver,n} + \varepsilon_{Car\ driver,n} \qquad (3.19)$$

$$U_{Shared\ Car,n} = \beta_t^{CM} * t_{Shared\ Car,n} + \varepsilon_{Shared\ Car,n} \qquad (3.20)$$

$$U_{Walking,n} = \beta_t^{W} * t_{Walking,n} + \varepsilon_{Walking,n} \qquad (3.21)$$

$$U_{Train,n} = \beta_c * c_{Train,n} + \beta_t^{PT} * t_{Train,n} + \beta_q * q_{Train,n} + \varepsilon_{Train,n} \qquad (3.22)$$

$$U_{Shared\ taxi,n} = \beta_c * c_{Shared\ taxi,n} + \beta_t^{PT} * t_{Shared\ taxi,n} + \beta_q * q_{Shared\ taxi,n} + \varepsilon_{Shared\ taxi,n} \qquad (3.23)$$

where cost (c_{in}) has a generic parameter (β_c) in the utility function of Bus, Car driver, Train and Shared taxi (i.e., we assumed that Shared Car and Walking had not cost); travel time (t_{in}), has a specific parameter depending on whether the mode was walking (β_t^W), Car driver and Shared car (β_t^{CM}) or public transport, that is Bus, Train and Shared taxi (β_t^{PT}), and the error term ε_{in} was considered independently and identically distributed (IID) Extreme Value Type I (0,1).

We considered that both c_{in} and t_{in}, were endogenous. We assumed that the endogeneity for c_{in} was due to measurement error and omitted variables, whereas the endogeneity for t_{in} was due to its simultaneous estimation in the supply-demand equilibration process. In this way we covered the three typical sources of endogeneity that may occur in the strategic transport modelling.

The variable c_{in} was constructed using the parameters shown in (3.24). Here, c_{in} is a function of an intercept term (α_c), q_{in}, the model's error term φ_{in}, and z_{in} which is an instrument required to correct for endogeneity in c_{in}.

$$c_{in} = \alpha_c + \gamma_z^c z_{in} + \gamma_q q_{in} + \varphi_{in} \qquad (3.24)$$

where γ_z^c is the parameter of z_{in}^c that considers the weakness/strength of the instrument in the simulation (Staiger and Stock, 1997) and should be much higher than zero. The parameter γ_q allows the correlation between c_{in} and q_{in}. If this value is zero, then endogeneity does not arise as will be explained below. Finally, the intercept term (α_c) is a constant that means a minimal cost.

On the other hand, a measurement error ω_{in} is added to the explanatory variable c_{in} as shown in (3.25), yielding the variable $\tilde{c}_{in}$ and, therefore, the first source of endogeneity for the simulation. In practice, this measurement error may come from the inaccuracy and/or complexity involved in the on-site data collection process, which is typical in large-scale mobility survey. Besides, the aggregation process during the modelling may also affect the estimation of the explanatory variables such as travel times and travel costs.

$$\tilde{c}_{in} = c_{in} + \omega_{in} \qquad (3.25)$$

The second source of endogeneity in our simulation comes from omitting q_{in} in (3.18), (3.19), (3.22) and (3.23), when the model is estimated using the simulated data. The inclusion of q_{in} in (3.24) accounts for this source of endogeneity in the simulation. Note that if q_{in} is omitted in the specification of the utility functions in (3.18), (3.19), (3.22) and (3.23), then the new error term in those equations would be equal to $\tilde{\varepsilon}_{in} = \beta_q q_{in} + \varepsilon_{in}$. Therefore $\tilde{c}_{in}$ would be correlated with $\tilde{\varepsilon}_{in}$, and by definition, $\tilde{c}_{in}$ would be endogenous. Note that this only affects the utility function for Bus, Car driver, Train and Shared taxi, because $\tilde{c}_{in}$ is an explanatory variable for these modes. The omission of q_{in} does not affect t_{in} because it is not correlated with q_{in}.

In practice, variables such as safety, comfort and/or reliability are often omitted in strategic transport models. These variables are usually significant in explaining the mode choice stage. Still, they are difficult to measure and can also be correlated with cost and/or travel time, causing additional endogeneity.

To create endogeneity in the variable t_{in}, we simulated a simultaneous equilibrium. For this, we reproduced the estimation process in the strategic transport modelling suite ESTRAUS[5] (De Cea *et al*., 2005), where the levels-of-service are estimated for the distribution, mode choice and assignment sub-models. First, six BPR (Bureau of Public Roads, 1964) type functions were defined, as shown in (3.26) to (3.31):

$$t_{Bus,n} = \pi_{Bus} * t_{Car\ driver,n} \tag{3.26}$$

$$t_{Car\ driver,n} = fft_{Car\ driver}\{1 + [\mu * demand(\tau_{Bus} * P_{Bus} + P_{Car\ driver} + \tau_{Shared\ taxi} * P_{Shared\ taxi})^{\rho}]\} \tag{3.27}$$

$$t_{Shared\ car,n} = t_{Car\ driver,n} \tag{3.28}$$

$$t_{Walking,n} = fft_{Walking,n} \tag{3.29}$$

$$t_{Train,n} = fft_{Train,n} \tag{3.30}$$

$$t_{Shared\ taxi,n} = \pi_{Shared\ taxi} * t_{Car\ driver,n} \tag{3.31}$$

where fft_{in} corresponds to the free-flow time for alternative *i*, which was considered a positive non-zero uniform random variable. These values are exogenous and come from previous research reported in the literature using real databanks (Guerrero *et al*., 2020). In practice, the parameters (μ and ρ) of the BPR function must be calibrated. We considered that Bus, Car driver, Shared car and Shared taxi used the same infrastructure; therefore, the travel times for Bus, Shared car and Shared taxi depend on the Car driver's travel time. In the same way, the Car driver's travel time is only affected by its choice probability ($P_{Car\ driver}$) and those of Bus and Shared taxi (P_{Bus} and $P_{Shared\ taxi}$). The parameters τ_{Bus} and $\tau_{Shared\ taxi}$ refer to the car equivalent units. On the other hand, the parameters π_{Bus} and $\pi_{Shared\ taxi}$ emulate increases in the travel time for the modes Bus and Shared taxi in comparison with the travel time for the car driver mode, given that they share the same infrastructure of transport. The variable called demand was added to the BPR function to consider the effect of increased demand over the years; it takes a positive non-zero value in the base year.

[5] ESTRAUS is a simultaneous equilibrium model designed to analyse and evaluate multimodal urban transport systems with multiple user classes, which has been extensively applied in Santiago and other Chilean cities.

Finally, given that Walking and Train do not share infrastructure with other modes, their travel times only depend on their free-flow time.

Having defined the travel time functions for each transport mode, the simulated simultaneous equilibrium for the *true* model followed the iterative process shown in Figure 3-1. The values reached at equilibrium will be known as $\tilde{t}_{in}$ from now on, and they are considered endogenous. First $\tilde{t}_{in}$ is estimated using the functions (3.26) to (3.31). Then, using the variable values obtained in (3.24) and the error term (ε_{in}), the utility functions in (3.18) to (3.23) are estimated. Finally, the probabilities obtained are used back again in step one, to find new values for $\tilde{t}_{in}$.

The iterative process must converge to an equilibrium, and the stop criterion is triggered when the tolerance (between $\tilde{t}_{in}$ in two successive iterations) is less than or equal to 0.001[6]. The convergence in our simulation was achieved implementing the method of successive weighted averages (MSWA) proposed by Liu *et al.* (2009). Note that fft_{in} is, by construction, a valid instrument of $\tilde{t}_{in}$; because fft_{in} is exogenous and correlated with $\tilde{t}_{in}$, as it appears in the right-hand side of equations (3.26) to (3.31).

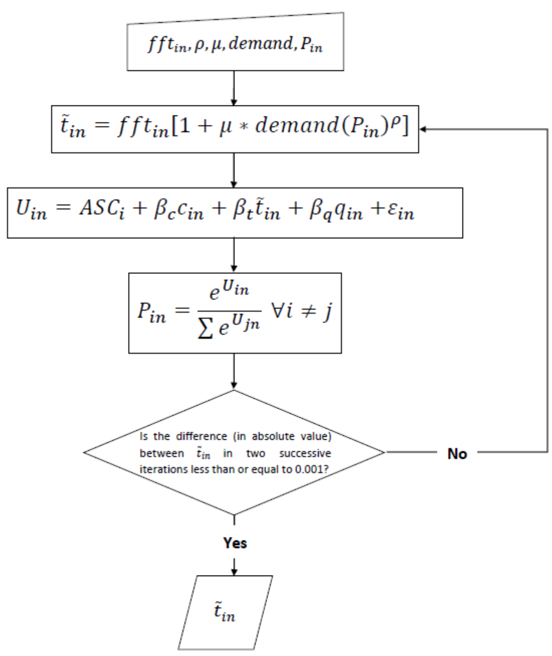

Figure 3-1. *Simulated simultaneous equilibrium process flowchart for the true model*

The choices for the base year were simulated with the parameters specified in Table 3-1.

[6] We made a sensitivity analysis considering changes in the tolerance (0.1, 0.01 and 0.0001). The results showed no significant differences with the tolerance initially contemplated.

Parameter	Value	Parameter	Value
β_t^{PT}	-0.03	π_{Bus}	1.25
β_t^{CM}	-0.04	$\pi_{Shared\ taxi}$	1.05
β_t^{W}	-0.05	μ	0.80
β_c	-0.01	ρ	1.20
α_c	10.0	τ_{Bus}	0.05
γ_z^c	2.0	$\tau_{Shared\ taxi}$	0.40
γ_q	1.0	*demand*	5

Table 3-1. *Parameters of the Monte Carlo simulation*

The values of the parameters β_t^{PT}, β_t^{CM}, β_t^{W} and β_c are typical of results found in applications of strategic urban mode choice models using DCM (Varela *et al*., 2018; Guerrero *et al*., 2020). The value of the parameter associated with γ_z^c represents the weakness/strength of the instrument in the simulation (Staiger and Stock, 1997). This value guarantees that the instrument is strong enough since it must be higher than zero. Besides, the value of γ_q associated with the variable q_{in} allows the correlation between c_{in} and q_{in}. If this value is zero, endogeneity does not arise as explained above. The intercept term (α_c) is a constant to guarantee that c_{in} will always be positive. The coefficients τ_{Bus} and $\tau_{Shared\ taxi}$ refer to the car equivalent units; for example, 0.05 comes from assuming that 20 people on a bus occupy the same road space as one person in a car (SECTRA, 2013). On the other hand, the parameters π_{Bus} and $\pi_{Shared\ taxi}$ mean that the travel times of Bus and Shared taxi are fixed as 25% and 5% higher than the Car driver's travel time, respectively (e.g., additional time spent to stop at stations to pick up passengers). Finally, μ and ρ are the BPR function parameters, which must be $\mu > 0$ and $\rho > 1$ (Márquez *et al*., 2014).

We estimated two different models for the base year: *endogenous* and *corrected*. The last one used the *CF* classical approach (Heckman 1978). The endogenous model includes $\tilde{c}_{in}$ and $\tilde{t}_{in}$ in the specification of the utility function, and the corrected model includes $\tilde{c}_{in}$, $\tilde{t}_{in}$, $\hat{\delta}_{in}^c$ and $\hat{\delta}_{in}^t$, where $\hat{\delta}_{in}^c$ and $\hat{\delta}_{in}^t$ are the residuals from the first stage of the *CF* approach for the endogenous variables $\tilde{c}_{in}$ and $\tilde{t}_{in}$, as shown in (3.32) and (3.33):

$$\tilde{t}_{in} = \alpha_t + \gamma_z^t z_{in} + \gamma_{fft}^t fft_{in} + \delta_{in}^t \quad \xrightarrow{OLS} \quad \hat{\delta}_{in}^t = \tilde{t}_{in} - t_{in} \quad (3.32)$$

$$\tilde{c}_{in} = \alpha_c + \gamma_z^c z_{in} + \gamma_{fft}^c fft_{in} + \delta_{in}^c \quad \xrightarrow{OLS} \quad \hat{\delta}_{in}^c = \tilde{c}_{in} - c_{in} \quad (3.33)$$

where α_t, α_c, γ_z^t, γ_z^c, γ_{fft}^t, γ_{fft}^c, $\hat{\delta}_{in}^t$ and $\hat{\delta}_{in}^c$ are obtained from an OLS regression of $\tilde{t}_{in}$ and $\tilde{c}_{in}$ on the exogenous variables z_{in} and fft_{in}. This is the first stage of the two-stage *CF* approach proposed by Wooldridge (2010).

Until here, this explains only the process of the choices' simulation and estimation of the endogenous and corrected model parameters. The forecasting stage has not started yet. To answer our research question - "how correcting for endogeneity would work in forecasting future scenarios (i.e., 10 to 40 years ahead) when the variables that change are endogenous?" - we compared three forecasting approaches: (*i*) do nothing (*No endogeneity correction*); (*ii*) the *Guevara and Ben-Akiva (2012) CF* approach and (*iii*) our proposal (*CFU* method). We considered two critical aspects of modelling transport in future scenarios: first, exogenous changes in the explanatory variables and second, increases in traffic demand (congestion). For the first aspect, we simulated several hypothetical scenarios where the travel time and/or travel cost change (for example, decreasing travel time due to infrastructure improvements or increasing travel cost due to change in fuel price).

In the simulation of future scenarios for each approach, fft_{in} and z_{in} can change in future situations. For example, infrastructure improvements may yield a decrease in free-flow times, identified as fft_{in}^{FUT} for future scenarios. On the other hand, increasing travel cost ($\tilde{c}_{in}^{FUT}$) can be due to the fuel price (represented in z_{in}^{FUT}) and suffer a shift for future scenarios; therefore, $\tilde{c}_{in}^{FUT}$ must be estimated as shown in (3.34):

$$\tilde{c}_{in}^{FUT} = \underbrace{\left(\alpha_c + \gamma_z^c z_{in}^{FUT} + \gamma_q q_{in} + \tilde{\delta}_{in}\right)}_{c_{in}^{FUT}} + \omega_{in} \qquad (3.34)$$

In particular, we simulated the effect of eight scenarios shown in Table 3-2.

Scenario	Exogenous change
1	50% decrease in free flow time for the train mode
2	30% decrease in free flow time for the train mode
3	10% decrease in free flow time for the car driver, shared car, bus and shared taxi modes
4	10% increase in travel cost for the car mode
5	50% decrease in free flow time for the train mode and 20% decrease in free flow time for the car driver, shared car, bus and shared taxi modes
6	50% decrease in free flow time for the train mode and 20% increase in travel cost for the car mode
7	30% decrease in free flow time for the train mode, 10% decrease in free flow time for the car driver, shared car, bus and shared taxi modes and 10% increase in travel cost for the car mode
8	50% decrease in free flow time for the train mode, 20% decrease in free flow time for the car driver, shared car, bus and shared taxi modes and 20% increase in travel cost for the car mode

Table 3-2. *Description of the hypothetical future scenarios*

Finally, to estimate the future value of the demand ($demand^{FUT}$), we applied a simple model (3.35):

$$demand^{FUT} = demand(1 + r)^N \qquad (3.35)$$

where *r* is the demand growth rate (say, 5%), and *N* is the forecasting year (i.e., 10 to 40 years).

In the case of the future scenarios, a new equilibrium must be reached as shown in Figure 3-2, which is affected by the changes in fft_{in}^{FUT}, $\tilde{c}_{i,n}^{FUT}$ and $demand^{FUT}$. The values reached in this equilibrium process are $\tilde{t}_{in}^{FUT}$, and are considered endogenous because they come from a simultaneous equilibrium process. Note that in the process described, U_{in}^{FUT} corresponds to the model utility corrected using the *CF* approach of Guevara and Ben-Akiva (2012) in the forecasting stage (i.e., (3.11) or (3.12)).

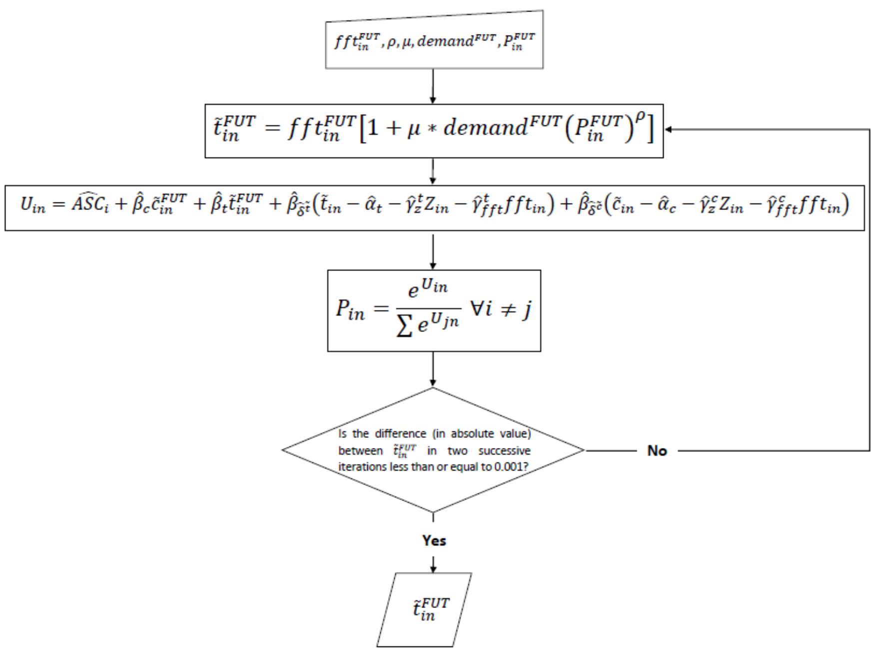

Figure 3-2. *Simultaneous equilibrium process for future scenarios using CF method*

Given that the equilibrium for future scenarios is reached in each approach, the specification of U_{in}^{FUT} must change, accordingly, in each approach. Besides, note that as we are forecasting, in general $\widehat{ASC}$, $\hat{\beta}$, $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\delta}$, the parameters used in U_{in}^{FUT} are those estimated in the base year by the *true*, endogenous, and corrected models. However, our new methodological approach (*CFU*) requires updating the residuals $\hat{\delta}$ as follows.

Once the values $\tilde{t}_{in}^{FUT}$ are estimated (Figure 3-2), we need to update the *CF* for the future scenarios as shown in (3.36) and (3.37). The update consists in replacing the instrument fft_{in} and z_{in} with fft_{in}^{FUT} and z_{in}^{FUT} (respectively) in the first stage of the *CFU* approach, so that new values for the residuals are obtained ($\hat{\delta}_{in}^{c\_FUT}$ and $\hat{\delta}_{in}^{t\_FUT}$).

$$\tilde{t}_{in}^{FUT} = \alpha_t + \gamma_z^t z_{in}^{FUT} + \gamma_{fft}^t fft_{in}^{FUT} + \delta_{in}^{\tilde{t}\_FUT} \quad \overset{OLS}{\longrightarrow} \quad \hat{\delta}_{in}^{\tilde{t}\_FUT} = \tilde{t}_{in}^{FUT} - t_{in}^{FUT} \quad (3.36)$$

$$\tilde{c}_{in}^{FUT} = \alpha_c + \gamma_z^c z_{in}^{FUT} + \gamma_{fft}^c fft_{in}^{FUT} + \delta_{in}^{\tilde{c}\_FUT} \quad \overset{OLS}{\longrightarrow} \quad \hat{\delta}_{in}^{\tilde{c}\_FUT} = \tilde{c}_{in}^{FUT} - c_{in}^{FUT} \quad (3.37)$$

For the second stage of the *CFU* approach, where $\hat{\delta}_{in}^{\tilde{t}\_FUT}$ and $\hat{\delta}_{in}^{\tilde{c}\_FUT}$ are added as explanatory variables within the utility function, we can re-estimate the parameters $\beta_{\hat{\delta}^{\tilde{t}\_FUT}}$ and $\beta_{\hat{\delta}^{\tilde{c}\_FUT}}$ using the Berkson-Theil transformation (Ortúzar and Willumsen 2011) as shown in (3.38) and (3.39). This can be done because all the terms (less $\beta_{\hat{\delta}^{\tilde{t}\_FUT}}$ and $\beta_{\hat{\delta}^{\tilde{c}\_FUT}}$) in (3.38) and (3.39) are scalars, as the modeller knows them.

$$Ln\frac{P_{in}^{FUT}}{P_{jn}^{FUT}} = \left(\widehat{ASC}_i - \widehat{ASC}_j\right) + \hat{\beta}_t\left(\tilde{t}_{in}^{FUT} - \tilde{t}_{jn}^{FUT}\right) + \hat{\beta}_c\left(\tilde{c}_{in}^{FUT} - \tilde{c}_{jn}^{FUT}\right) + \beta_{\hat{\delta}\tilde{t}\_FUT}\left(\hat{\delta}_{in}^{\tilde{t}\_FUT} - \hat{\delta}_{jn}^{\tilde{t}\_FUT}\right) + \beta_{\hat{\delta}\tilde{c}\_FUT}\left(\hat{\delta}_{in}^{\tilde{c}\_FUT} - \hat{\delta}_{jn}^{\tilde{c}\_FUT}\right) \quad (3.38)$$

$$Ln\frac{P_{in}^{FUT}}{P_{jn}^{FUT}} - \left\{\left(\widehat{ASC}_i - \widehat{ASC}_j\right) + \hat{\beta}_t\left(\tilde{t}_{in}^{FUT} - \tilde{t}_{jn}^{FUT}\right) + \hat{\beta}_c\left(\tilde{c}_{in}^{FUT} - \tilde{c}_{jn}^{FUT}\right)\right\} = \beta_{\hat{\delta}\tilde{t}\_FUT}\left(\hat{\delta}_{in}^{\tilde{t}\_FUT} - \hat{\delta}_{jn}^{\tilde{t}\_FUT}\right) + \beta_{\hat{\delta}\tilde{c}\_FUT}\left(\hat{\delta}_{in}^{\tilde{c}\_FUT} - \hat{\delta}_{jn}^{\tilde{c}\_FUT}\right) \quad (3.39)$$

In this way, $\beta_{\hat{\delta}\tilde{t}\_FUT}$ and $\beta_{\hat{\delta}\tilde{c}\_FUT}$ can be re-estimated (i.e. updated) using an OLS regression, and a new equilibrium point is reached considering the updated parameters $\beta_{\hat{\delta}\tilde{t}\_FUT}$ and $\beta_{\hat{\delta}\tilde{c}\_FUT}$ (Figure 3-3). The values reached in this new equilibrium are $\tilde{\tilde{t}}_{in}^{FUT}$.

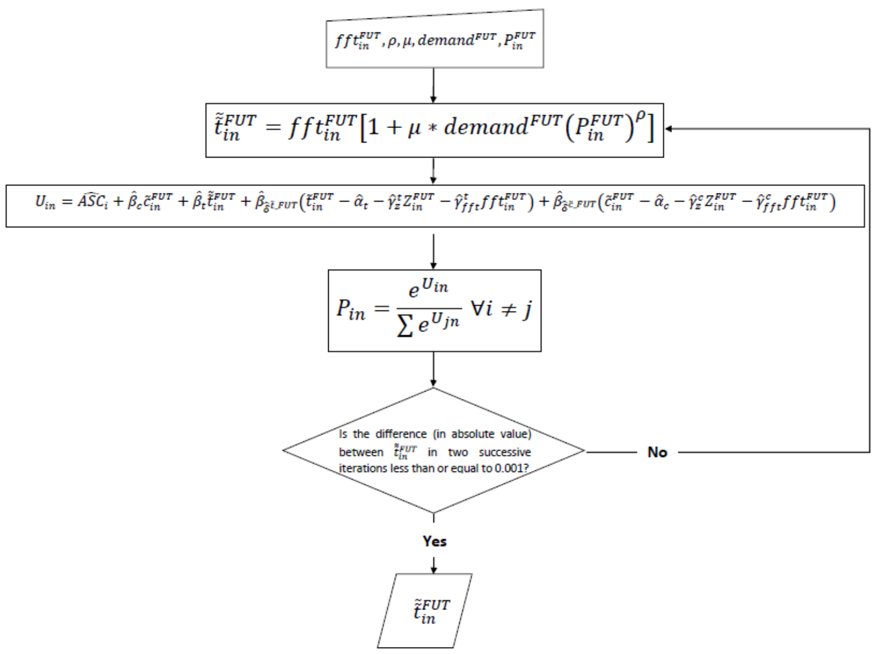

Figure 3-3. *Simulated simultaneous equilibrium process flowchart for future scenarios with the CFU approach*

3.4. Results

3.4.1. *Base year assessment*

As the correction for endogeneity in a DCM produces a change of scale in the parameter estimates, we checked the ratios among parameters. Let β_t^{PT} be the parameter of travel time for the public transport modes (Bus, Train and Share Taxi), β_t^{CM} for the car modes (Car driver and Share car), β_t^{W} for Walking and β_c the parameter of cost (which was considered generic). Using the parameters shown in Table 3-1, we can see that the *true* population ratios are $\frac{\beta_t^{PT}}{\beta_c} = \frac{-0.03}{-0.01} = 3$, $\frac{\beta_t^{W}}{\beta_c} = \frac{-0.05}{-0.01} = 5$ and $\frac{\beta_t^{CM}}{\beta_c} = \frac{-0.04}{-0.01} = 4$. Our results for the benchmark (*true* model), endogenous and corrected models are reported in Table 3-3, which shows the mean of the ratios $\frac{\beta_t^{PT}}{\beta_c}$, $\frac{\beta_t^{W}}{\beta_c}$ and $\frac{\beta_t^{CM}}{\beta_c}$, the t-test against the *true* ratios and the bias (in percentage) for each case. For the base year, only the classical *CF* approach is used to correct for endogeneity and to recover the parameters. Our approach (*CFU*) is applied later for forecasting.

Model	$\frac{\beta_t^{PT}}{\beta_c}$ (t-test)	$\frac{\beta_t^{W}}{\beta_c}$ (t-test)	$\frac{\beta_t^{CM}}{\beta_c}$ (t-test)	**% Bias** $\frac{\beta_t^{PT}}{\beta_c}$	**% Bias** $\frac{\beta_t^{W}}{\beta_c}$	**% Bias** $\frac{\beta_t^{CM}}{\beta_c}$
True	3.00	5.00	4.00	-	-	-
Endogenous	1.47 (26.89)	10.19 (13.78)	3.67 (3.36)	51.1%	103.8%	8.3%
Corrected	2.97 (0.49)	4.92 (0.58)	3.94 (1.03)	0.9%	1.6%	1.6%

Table 3-3. *Statistics for benchmark, corrected and endogenous model*

The biases for the endogenous ratios are large, varying from 8.3% (i.e., $\left[\frac{4.00-3.67}{4.00}\right] * 100$) to 103.8% ($\left[\frac{10.19-5.00}{5.00}\right] * 100$). In this case too, the t-test[7] against the null hypotheses that $\frac{\beta_t^{PT}}{\beta_c} = 3$, $\frac{\beta_t^{W}}{\beta_c} = 5$ and $\frac{\beta_t^{CM}}{\beta_c} = 4$ can be easily rejected and it can be concluded that the parameter ratios in the endogenous model are significantly different from the *true* values for a one-sided test at the 95% confidence level[8]. On the other hand, the t-test for the parameter ratios in the corrected model are all accepted for a one-sided test at the 95% confidence level, meaning that the corrected ratios are not significantly different from the *true* ratios. The biases for the corrected ratios are small, varying from 0.9% to 1.6%.

The boxplots[9] in Figure 3-4 show the parameter ratios for the endogenous and corrected model using the classical *CF* approach (Heckman 1978). We do not show the median; instead, we show the mean as a black dot depicting the average for the 100 repetitions, as we are interested in the average of all observations. The dashed line represents the *true* ratio value. The variance of the corrected ratios seems lower than those of the endogenous ratios. Also, there are more outliers for the endogenous ratios than for the corrected ratios. The former is far from the *true* ratios, showing the impact of endogeneity in yielding inconsistent parameters. The parameter ratios for the corrected model are close to the *true* ratios, showing the power of the *CF* method to correct for endogeneity.

[7] The t-test statistic has the form $t = \frac{\bar{X}-\mu}{sd/\sqrt{n}}$ where $\bar{X}$ is the sample mean from a sample $X_1, X_2, \ldots, X_n$, of size n, sd is the estimate of the standard deviation of the population, and μ is the population mean.
[8] When the sign of the parameter is known a one-sided test should be applied; the critical value of t is 1.64 for a one-sided test at the 95% confidence level.
[9] Boxplots were introduced by Tukey (1977). They consist of a rectangle with bottom and topside at the 1st and 3rd quartile (i.e., 25th and 75th percentiles). The distance between the 1st and 3rd quartiles is known as the *interquartile range* (IQR). It allows getting an idea of the dispersion (accumulation) of the values drawn. Usually, they have a horizontal line added at the median (2nd quartile or 50th percentile), and other lines known as *whiskers*. They have a length 1.5 times the IQR added at the top and bottom. Observations outside of the whiskers are plotted and considered as *outliers*.

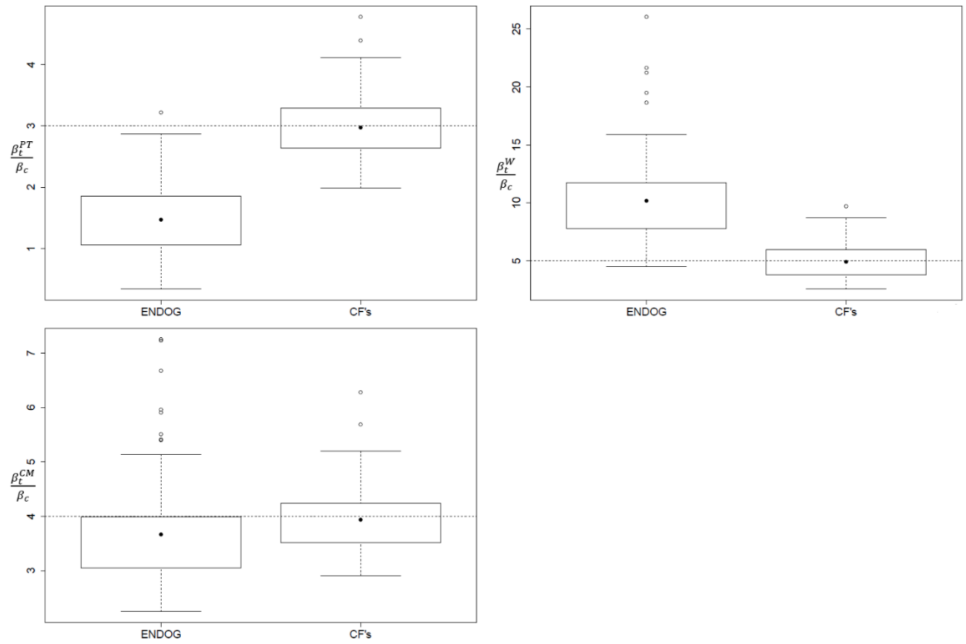

Figure 3-4. *Boxplots of parameter ratios for the endogenous and CF corrected model*

Note that for all the cases analysed, reported in Table 3-4, the TTE for the endogenous model were larger than those of the true and corrected models. This does not need to be the case in general. In our case study, the TTE for the endogenous model was larger because, for the specific settings considered, the endogenous model resulted in an underestimation of the value of time of the motorized modes that experienced congestion (see Figure 3-4). So, smaller values of time implied that the simulated individuals in that case were willing to pay less for reducing travel time (compared to what they would do in the true model), resulting in choices of slower alternatives and, hence, larger TTE.

3.4.2. *Future scenarios assessment*

Exogenous changes impact individual choices. This happens because the exogenous changes affect the explanatory variables of the model (travel time and cost), which in turn modify the supply-demand equilibration in future scenarios. With our Monte Carlo simulation, we estimate several measures (travel times at equilibrium, the logarithm of the likelihood expected value, and the expected value of the Akaike Information Criteria) and compare them with those obtained for the *true* model.

Table 3-4. *Average free flow time and travel time in equilibrium (TTE) for 100 replications*

Year	Approach	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
		Car modes	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**
Calibration	Free flow time	45.03	49.53	35	45.03	49.53	35	45.03	49.53	35	45.03	49.53	35
	TTE	64.04	64.1	35	64.04	64.1	35	64.04	64.1	35	64.04	64.1	35
10	TTE True	72.28	62.91	35	73.2	66.62	35	68.74	67.7	35	73.29	71.19	35
	TTE No endogeneity correction	75.61	65.47	35	75.89	68.68	35	69.27	68.11	35	75.52	72.9	35
	TTE Guevara and Ben-Akiva (2012)	71.79	62.54	35	72.66	66.21	35	67.79	66.97	35	72.68	70.72	35
	TTE *CFU*	71.51	62.33	35	72.52	66.1	35	68.06	67.18	35	72.76	70.78	35
20	TTE True	87.78	74.8	35	89	78.73	35	83.91	79.33	35	89.19	83.38	35
	TTE No endogeneity correction	97.5	82.25	35	97.95	85.59	35	89.68	83.75	35	97.51	89.75	35
	TTE Guevara and Ben-Akiva (2012)	89.3	75.96	35	90.54	79.91	35	84.89	80.08	35	90.62	84.47	35
	TTE *CFU*	87.98	74.95	35	89.35	79	35	84.26	79.6	35	89.69	83.76	35
30	TTE True	115.63	96.15	35	117.15	100.32	35	109.68	99.09	35	117.5	105.08	35
	TTE No endogeneity correction	135.35	111.27	35	136.04	114.8	35	124.79	110.67	35	135.7	119.04	35
	TTE Guevara and Ben-Akiva (2012)	122.39	101.33	35	124.02	105.58	35	115.22	103.34	35	124.31	110.3	35
	TTE *CFU*	119.97	99.48	35	121.64	103.75	35	113.12	101.72	35	122.04	108.56	35
40	TTE True	170.61	138.3	35	172.41	142.68	35	159.6	137.36	35	173.05	147.67	35
	TTE No endogeneity correction	200.12	160.92	35	201.11	164.68	35	184.73	156.63	35	201.11	169.18	35
	TTE Guevara and Ben-Akiva (2012)	185.76	149.92	35	187.74	154.44	35	173.91	148.33	35	188.5	159.51	35
	TTE *CFU*	183.38	148.09	35	185.35	152.6	35	171.48	146.47	35	186.11	157.68	35
Free flow time with exogenous change		45.03	42.03	35	45.03	45.03	35	40.53	46.07	35	45.03	49.53	35

Year	Approach	Scenario 5			Scenario 6			Scenario 7			Scenario 8		
		Car modes	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**
Calibration	Free flow time	45.03	49.53	35	45.03	49.53	35	45.03	49.53	35	45.03	49.53	35
	TTE	64.04	64.1	35	64.04	64.1	35	64.04	64.1	35	64.04	64.1	35
10	TTE True	60.66	54.01	35	69.97	61.14	35	66.20	61.26	35	58.44	52.30	35
	TTE No endogeneity correction	61.57	54.7	35	74.10	64.31	35	68.14	62.74	35	60.21	53.66	35
	TTE Guevara and Ben-Akiva (2012)	59.59	53.19	35	69.51	60.79	35	65.40	60.64	35	57.48	51.57	35
	TTE *CFU*	59.78	53.33	35	69.29	60.63	35	65.50	60.72	35	57.69	51.73	35
20	TTE True	74.03	64.26	35	84.78	72.50	35	80.55	72.26	35	71.11	62.01	35
	TTE No endogeneity correction	79.69	68.59	35	95.35	80.60	35	87.98	77.95	35	77.68	67.05	35
	TTE Guevara and Ben-Akiva (2012)	74.52	64.63	35	86.08	73.49	35	81.46	72.96	35	71.48	62.30	35
	TTE *CFU*	73.88	64.14	35	84.86	72.56	35	80.66	72.34	35	70.93	61.88	35
30	TTE True	96.34	81.36	35	112.05	93.40	35	105.55	91.42	35	92.76	78.62	35
	TTE No endogeneity correction	110.83	92.47	35	132.68	109.22	35	122.46	104.39	35	108.2	90.45	35
	TTE Guevara and Ben-Akiva (2012)	100.37	84.45	35	118.42	98.29	35	110.77	95.43	35	96.39	81.40	35
	TTE *CFU*	98.34	82.9	35	116.07	96.49	35	108.58	93.75	35	94.45	79.91	35
40	TTE True	139.51	114.45	35	166.77	135.36	35	154.85	129.22	35	135.52	111.4	35
	TTE No endogeneity correction	164.36	133.51	35	197.21	158.69	35	181.82	149.90	35	161.39	131.23	35
	TTE Guevara and Ben-Akiva (2012)	151.9	123.95	35	181.64	146.76	35	168.82	139.93	35	147.61	120.67	35
	TTE *CFU*	149.44	122.07	35	179.31	144.97	35	166.42	138.09	35	145.22	118.83	35
Free flow time with exogenous change		36.03	35.12	35	45.03	42.03	35	40.53	41.57	35	40.53	36.03	35

We assessed and compared the new CFU method with (1) *No endogeneity correction* and (2) the *Guevara and Ben-Akiva (2012) CF* approach for the eight scenarios shown in Table 3-2. In each case, the model forecasts were evaluated in terms of their ability to recover the *true* (simulated) travel times at equilibrium (TTE), the logarithm of the likelihood expected value - $E(l(\theta))$ for the *true* model, and the expected value of the Akaike Information Criteria[10] – $E(AIC)$ for future scenarios in 10, 20, 30 and 40 years ahead. The *true* model is the benchmark and $E(l(\theta))$ and $E(AIC)$ were calculated as shown in (3.40) and (3.41), respectively:

$$E(l(\theta)_I^{m,t}) = \sum_n \sum_{i \epsilon A(n)} ln\left(P_{in}^{m,t} * P_{in}^{True,t}\right) \qquad (3.\ 40)$$

$$E(AIC_I^{m,t}) = 2k - 2E(l(\theta)_I^{m,t}) \qquad (3.\ 41)$$

where $P_{in}^{True,t}$ represents the choice probability of alternative *i* for individual *n* in the *true* model for year *t*; $P_{in}^{m,t}$ is the choice probability of alternative *i* for individual *n* in approach *m* (*No endogeneity correction*, *Guevara and Ben-Akiva (2012)* and *CFU*) for year *t*, and the subscript (*I*) indicates the number of repetitions. We used $E(l(\theta)_I^{m,t})$ to estimate $E(AIC_I^{m,t})$ for approach *m* and year *t,* where *k* is the number of model parameters.

Table 3-4 summarizes the average for all repetitions of the free flow time in the base year and with the exogenous change (see Table 3-2), as well as the travel times achieved for the future equilibria in the simulations for each scenario. Results are shown for the base and forecasting years for the three approaches, as well as for the benchmark (*true* model). As can be seen, for the base year, the TTE are the same in all scenarios because the estimates replicate the values. Note that the free flow time and the TTE do not change for the Walking mode, because walking does not share infrastructure with other modes; therefore, it is not affected by congestion. The TTE (in the base year) are higher than the free flow time because of congestion[11].

To evaluate the future scenarios in terms of recovering the *true* (simulated) travel times, we used the % bias as shown in (3.42). This way, we can indicate how well or poorly the TTE are recovered in comparison with the *true* model. The TTE for the 10-to-40-year forecasts with the endogenous model are worse than in the *true* model in all scenarios. Also, the effects of "*No endogeneity correction*" increase with time.

$$\%\ bias = \frac{(\text{TTE}_{Approach} - \text{TTE}_{true})}{\text{TTE}_{true}} * 100 \qquad (3.\ 42)$$

[10] AIC is a measure of the relative quality of a statistical model. It provides a trade-off between the goodness of fit of the model and its complexity (Akaike, 1974).

[11] We simulated a case with endogeneity only in travel time due to the equilibrium conditions (i.e., no endogeneity due to measurement error and omitted variables), so cost was not endogenous. We found that the effect was similar but smaller than in the case with three sources of endogeneity, an expected result; as there is less bias, the TTE will tend to be closer to those in the true model. So, there appears to be an additive effect regarding endogeneity sources; that is, if the endogeneity sources increase, the bias also does.

On the other hand, both the *CFU* and *Guevara and Ben-Akiva (2012)* results are substantially better than those of the endogenous model, but they are still different from the *true* model. This can be attributed to a simulation error. Notwithstanding, the TTE reached with the *CFU* approach are closer to the values of the *true model* than the TTE reached with the *Guevara and Ben-Akiva (2012)* approach.

We applied the t^*-statistics (Ortúzar and Willumsen 2011, 341-342) to test the null hypothesis that the mean difference of TTE for the 100 repetitions between *CFU* and *Guevara and Ben-Akiva (2012)* was zero. The t^*-statistics applied for 20, 30 and 40 years show that these values are superior to the critical value (i.e., 1.96 for a two-sided test at the 95% confidence level); therefore, the null hypothesis is rejected and there is a significant difference between the means. Consequently, we can conclude that the *CFU* approach is better than the approach proposed by *Guevara and Ben-Akiva (2012)*. This happened in all the scenarios analysed.

Figure 3-5 focuses only on the car modes and shows the boxplot for the TTE reached by each of the three approaches. Again, the *true* model's travel times are the benchmark. These are represented by the dashed line drawn at zero. The points that belong to the boxplots are the difference between the TTE reached from any other approaches and the *true* model. We show only the case corresponding to Scenario 8 for explanatory purposes because it is one of two scenarios affected by the most significant number of exogenous shifts. Besides, for the other scenarios, the performance was very similar. As can be seen, "*No endogeneity correction*" is the worst in recovering the TTE. Figure 3-5 also shows that the endogenous model overestimates congestion. This makes sense given that, as shown in Table 3-3, the parameter ratios show large bias for the endogenous estimations.

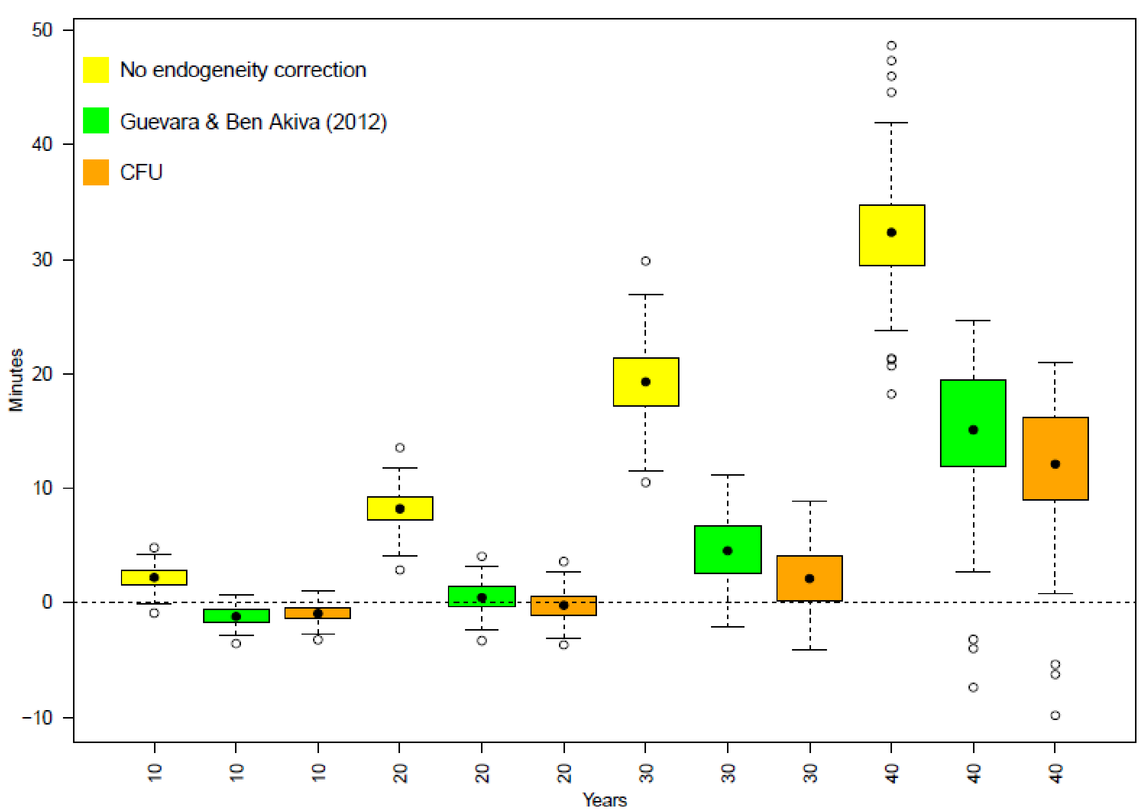

Figure 3-5. *TTE reached with the three approaches compared vs true model for Scenario 8*

To assess the relative quality of each approach, we used again $E(l(\theta))$ and $E(AIC)$. A summary of their estimates for each approach and year are given in Table 3-5 for the scenarios shown in Table 3-2, and all the repetitions run in the simulation.

Given that "*No endogeneity correction*" uses a model (endogenous) that is a restricted version of the model used in both *Guevara and Ben-Akiva (2012)* and in the *CFU* approach, it is possible to apply the likelihood ratio (LR) test (Ortúzar and Willumsen 2011, 281) to compare them. LR is asymptotically distributed χ^2_r with *r* degrees of freedom, where *r* is the number of linear restrictions required to transform the more general model into the restricted version. The null hypothesis is that the two models compared are equivalent; rejecting it, implies that the restricted model is erroneous. In our case, *r* = 2 (because the restrictions are that both $\hat{\beta}_{\widehat{\delta}\tilde{t}}$ and $\hat{\beta}_{\widehat{\delta}\tilde{c}}$ are zero).

Table 3-5. *Average estimates of E(l(θ)) and E(AIC) for 100 replications*

Year	Approach	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6		Scenario 7		Scenario 8	
		$E(l(\theta))$	$E(AIC)$	$E(l(\theta))$	$E(AIC)$	$E(l(\theta))$	$E(AIC)$	$E(l(\theta))$	$E(AIC)$	$E(l(\theta))$	$E(AIC)$	$E(l(\theta))$	$E(AIC)$	$E(l(\theta))$	$E(AIC)$	$E(l(\theta))$	$E(AIC)$
10	No endogeneity correction	-5303.4	10624.8	-5296.3	10610.6	-5378.5	10775.0	-5259.3	10536.7	-5508.2	11034.4	-5265.8	10549.6	-5375.7	10769.5	-5459.9	10937.9
	Guevara and Ben-Akiva (2012)	-5194.3	10410.5	-5212.2	10446.4	-5336.7	10695.4	-5183.2	10388.4	-5482.2	10986.3	-5157.6	10337.2	-5329.3	10680.6	-5427.4	10876.9
	CFU	-5136.0	10293.9	-5157.9	10337.9	-5298.2	10618.4	-5135.1	10292.2	-5449.2	10920.4	-5102.9	10227.8	-5288.3	10598.7	-5398.5	10818.9
20	No endogeneity correction	-5052.8	10123.6	-5046.4	10110.8	-5149.0	10316.0	-5024.1	10066.1	-5292.5	10603.0	-5037.7	10093.4	-5154.8	10327.6	-5266.0	10549.9
	Guevara and Ben-Akiva (2012)	-4871.9	9765.7	-4892.2	9806.4	-5056.0	10134.0	-4890.0	9802.0	-5215.5	10453.0	-4869.2	9760.5	-5055.5	10133.0	-5195.3	10412.6
	CFU	-4785.4	9592.9	-4807.2	9636.5	-4978.9	9979.7	-4807.7	9637.4	-5141.1	10304.3	-4783.7	9589.5	-4976.5	9975.0	-5123.9	10269.9
30	No endogeneity correction	-4694.9	9407.8	-4687.8	9393.6	-4814.0	9646.0	-4679.9	9377.8	-4980.3	9978.6	-4706.1	9430.2	-4832.4	9682.9	-4980.8	9979.6
	Guevara and Ben-Akiva (2012)	-4402.0	8826.0	-4417.0	8856.0	-4606.5	9235.1	-4430.8	8883.5	-4797.7	9617.3	-4434.3	8890.5	-4625.2	9272.4	-4816.8	9655.6
	CFU	-4319.1	8660.3	-4333.0	8687.9	-4517.3	9056.5	-4343.8	8709.6	-4706.6	9435.3	-4348.0	8718.1	-4534.9	9091.9	-4724.2	9470.4
40	No endogeneity correction	-4191.0	8400.0	-4180.7	8379.3	-4327.8	8673.5	-4180.4	8378.7	-4529.2	9076.5	-4222.9	8463.7	-4361.6	8741.2	-4556.4	9130.7
	Guevara and Ben-Akiva (2012)	-3793.1	7608.2	-3796.3	7614.6	-3982.1	7986.1	-3808.1	7638.2	-4213.3	8448.5	-3847.5	7717.0	-4027.3	8076.7	-4264.8	8551.6
	CFU	-3741.0	7504.0	-3742.6	7507.1	-3918.1	7858.3	-3749.9	7521.8	-4143.7	8309.3	-3791.0	7603.9	-3963.0	7947.9	-4190.5	8403.1

The LR test for the results of scenario 8 and year 10[12] comparing "*No endogeneity correction*" against *Guevara and Ben-Akiva (2012)* is $LR = -2(-5459.9 + 5427.4) = 65$, and comparing "*No endogeneity correction*" against *CFU* is $LR = -2(-5459.9 + -5398.5) = 122.8$. These values must be compared with the critical value for two degrees of freedom at the 95% confidence level ($\chi_2^2 = 5.99$). As LR > χ_2^2 (for both cases), the null hypothesis is confidently rejected, and we can conclude that the corrected version models are superior. Given that the *CFU* approach has $E(l(\theta))$ better than the *Guevara and Ben-Akiva (2012)* approach, we can conclude that the *CFU* approach performs best. Given that $E(AIC)$ values are calculated using the $E(l(\theta))$ and *k* (see expression 3.41), then it is expected that the $E(AIC)$ from *CFU* approach will also be better than those of the other methods (see Table 3-5).

Figure 3-6 deploys the boxplot of $E(l(\theta))$ estimations for the three approaches above, across the 100 repetitions of the simulation for Scenario 8. As can be seen, "*No endogeneity correction*" shows the worst $E(l(\theta))$. These findings reinforce the negative impact of endogeneity in forecasting. Besides, the *CFU* approach shows, once more, a better $E(l(\theta))$ over the years in comparison with the *Guevara and Ben-Akiva (2012)* approach. These findings confirm the conclusion that suggests *CFU* shows better performance over *Guevara and Ben-Akiva (2012)*, both of which are far from the endogenous model.

Finally, we show the average for 100 replications of the base year's market shares and the future scenarios in Table 3-6. Note that ignoring the impact of endogeneity does not affect the market shares for the base year, despite the significant bias in the model estimates, because the calibration of a strategic transport model always implies replicating the base year results. Note that the market shares estimated from *Guevara and Ben-Akiva (2012)* and *CFU* are consistently similar. In most cases, the market shares estimated with the "No endogeneity correction" approach are far worse than those calculated with the Guevara and Ben-Akiva (2012), and CFU approaches, except for two cases (in Scenario 1 for years 30 and 40).

[12] For the other scenarios and years, the performance is similar.

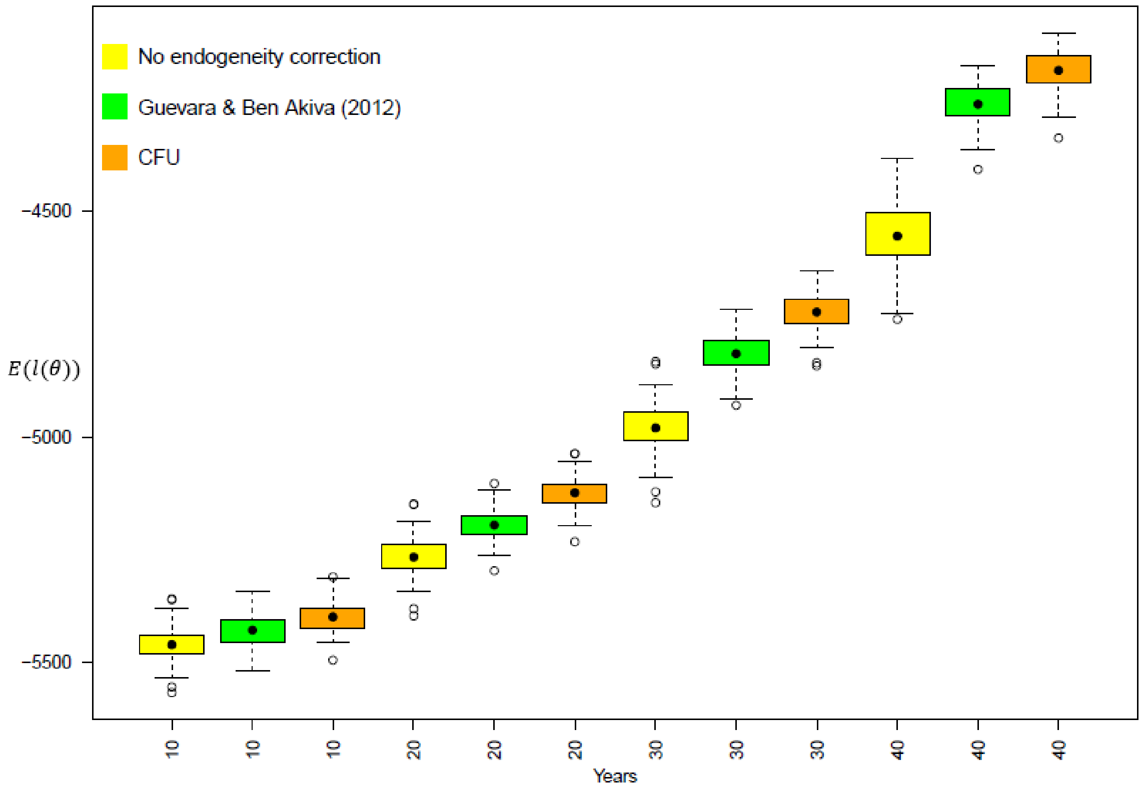

Figure 3-6. *Boxplots of $E(l(\theta))$ for approaches over the years for scenario 8*

As the results in Table 3-6 are shown as aggregate probabilities by mode, namely, Public modes (Bus, Train and Shared Taxi), Car modes (Car driver and Shared car), and Walking mode, this could be purely circumstantial. Our primary interest was to ensure that the CFU approach performed better on measures such as TTE, the logarithm of the expected probability value, and the expected value of the Akaike Information Criteria.

3.5. Conclusions and Future Research Directions

This chapter provides a framework for correcting endogeneity, and the correction applied at the forecasting stage of supply-demand equilibration models when the endogenous variables change over the years. We emulated a complex transport modelling process affected by three different endogeneity sources that are common in strategic studies (measurement error, omitted variables and simultaneous estimation in a supply-demand equilibration context). In this setting, we compared three different approaches: (1) No endogeneity correction (which has been, so far, the only method used in practice), (2) the *CF* approach proposed by Guevara and Ben-Akiva (2012) and (3) our new proposal, the *CFU* approach. Forecasts were evaluated in terms of recovery of the *true* (simulated) travel times, and two goodness-of-fit indices, $E(l(\theta))$ and the $E(AIC)$, for future scenarios in 10 to 40 years ahead.

Table 3-6. *Average of the market shares for 100 replications*

Year	Approach	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
		Car modes	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**
Calibration	Endogenous	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0
	Corrected	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0
10	True	26.6	48.8	24.6	27.3	47.3	25.4	29.9	44.8	25.3	28.3	45.4	26.3
	No endogeneity correction	28.4	46.4	25.3	28.6	46.1	25.4	30.0	44.9	25.2	28.7	45.6	25.6
	Guevara and Ben-Akiva (2012)	26.7	49.0	24.3	27.4	47.5	25.1	29.9	44.9	25.2	28.4	45.6	26.1
	CFU	26.4	49.1	24.5	27.2	47.5	25.3	29.9	44.7	25.4	28.2	45.5	26.3
20	True	23.9	50.1	26.0	24.5	48.7	26.8	27.0	46.1	26.9	25.5	46.7	27.8
	No endogeneity correction	25.2	48.7	26.1	25.4	48.4	26.3	26.9	47.0	26.0	25.6	47.9	26.5
	Guevara and Ben-Akiva (2012)	23.5	50.7	25.8	24.2	49.3	26.6	26.6	46.6	26.8	25.1	47.3	27.6
	CFU	23.2	50.9	26.0	23.8	49.3	26.8	26.4	46.5	27.1	24.8	47.2	27.9
30	True	20.5	52.0	27.5	21.1	50.5	28.4	23.4	47.9	28.7	22.1	48.5	29.5
	No endogeneity correction	20.7	52.0	27.3	20.8	51.7	27.5	22.4	50.3	27.3	21.0	51.3	27.7
	Guevara and Ben-Akiva (2012)	19.2	53.4	27.5	19.7	52.0	28.4	22.0	49.2	28.8	20.5	50.0	29.5
	CFU	19.1	53.5	27.5	19.6	52.0	28.4	21.9	49.2	29.0	20.5	50.0	29.6
40	True	16.8	54.1	29.1	17.3	52.7	30.0	19.3	50.1	30.6	18.2	50.6	31.2
	No endogeneity correction	15.6	55.5	28.9	15.7	55.2	29.1	17.1	54.0	28.9	15.9	54.8	29.3
	Guevara and Ben-Akiva (2012)	14.3	56.4	29.3	14.7	55.0	30.3	16.6	52.4	31.0	15.4	53.1	31.5
	CFU	14.5	56.4	29.1	14.9	55.0	30.2	16.7	52.4	30.9	15.6	53.0	31.4

Year	Approach	Scenario 5			Scenario 6			Scenario 7			Scenario 8		
		Car modes	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**	**Car modes**	**Public modes**	**Walking mode**
Calibration	Endogenous	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0
	Corrected	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0	30.7	44.3	25.0
10	True	29.7	47.7	22.6	26.4	49.1	24.5	28.8	47.0	24.3	29.5	48.0	22.5
	No endogeneity correction	30.6	45.0	24.4	28.2	46.5	25.3	29.6	45.5	25.0	30.4	45.2	24.4
	Guevara and Ben-Akiva (2012)	29.6	47.8	22.6	26.5	49.3	24.2	28.8	47.1	24.1	29.4	48.2	22.5
	CFU	29.5	47.8	22.6	26.2	49.4	24.4	28.6	47.1	24.3	29.3	48.2	22.5
20	True	27.0	48.9	24.1	24.0	50.2	25.8	26.1	48.1	25.8	27.0	49.0	23.9
	No endogeneity correction	27.9	47.0	25.2	25.2	48.7	26.1	26.7	47.5	25.8	27.8	47.0	25.2
	Guevara and Ben-Akiva (2012)	26.7	49.3	24.0	23.6	50.8	25.6	25.7	48.6	25.6	26.7	49.4	23.9
	CFU	26.4	49.4	24.2	23.3	51.0	25.8	25.4	48.7	25.9	26.4	49.6	24.0
30	True	23.6	50.6	25.8	20.8	51.9	27.3	22.8	49.8	27.4	23.9	50.5	25.6
	No endogeneity correction	23.6	50.1	26.3	20.7	52.0	27.3	22.3	50.8	27.0	23.7	50.0	26.3
	Guevara and Ben-Akiva (2012)	22.5	51.7	25.9	19.4	53.4	27.2	21.4	51.2	27.5	22.7	51.7	25.6
	CFU	22.3	51.8	26.0	19.4	53.4	27.2	21.3	51.2	27.6	22.5	51.8	25.7
40	True	19.7	52.8	27.6	17.3	53.9	28.8	18.9	52.0	29.1	20.1	52.6	27.3
	No endogeneity correction	18.3	53.8	27.9	15.8	55.4	28.8	17.1	54.4	28.6	18.5	53.7	27.8
	Guevara and Ben-Akiva (2012)	17.2	54.9	27.9	14.7	56.3	29.0	16.3	54.3	29.5	17.6	54.8	27.6
	CFU	17.3	54.9	27.8	14.8	56.3	28.9	16.4	54.3	29.4	17.7	54.8	27.5

This was a challenge, because this type of modelling is usually done using commercial software packages, where the level-of-service variables are estimated from complex supply-demand equilibria processes considering multiple user classes, several transport modes, complex networks and many other aspects.

Monte Carlo simulation helped us to demonstrate that under fairly reasonable conditions, consistent with observed data, the new *CFU* proposal performs better than both other approaches; in particular, it performs much better than the "*doing nothing*" approach and marginally (but significantly) better than the more classical *CF* approach of Guevara and Ben-Akiva (2012). We also show that the adverse effects of endogeneity increase over the years severely impacting the forecasts for future scenarios.

Our methodological findings suggest two important recommendations for practice. The first is to avoid forecasting with endogenous models as the problem is severe in the case studied. The second is that even when correcting for endogeneity, in forecasting the residuals from the first stage of the *CF* approach for the future scenarios should always be updated. Thus, our new CFU approach is especially recommended to correct for endogeneity when discrete choice models are used to forecast strategic scenarios involving supply-demand equilibration.

Three areas for further research can be identified. First, we believe it is crucial to examine in greater depth how the social evaluation of transport projects may be affected by endogeneity. Especially given our findings regarding changes in the level-of-service variables when forecasting over several years. Second, the functional form used in our simulation was of multinomial logit type. We recommend exploring other functional forms, such as Nested Logit or Mixed Logit, to see if results vary (although the latter is certainly difficult to implement in the context of a large-scale supply-demand equilibration model). Finally, an exciting extension of this research would be to apply our methodological proposal to real data.

4. A Monte Carlo Method to Detect Weak Instruments: Application to Linear and Discrete Choice Models

4.1. Introduction

This chapter proposes an alternative empirical approach to detect weak instruments in linear models and DCM, using Monte Carlo simulation. For this, we extend and adapt the Monte Carlo methodology proposed by Guevara and Navarro (2015), which is based on the criteria of *relative bias* (RB) proposed by Stock and Yogo (2005). We will first validate our empirical approach, in the case of linear models, by comparing its results with critical values derived analytically (Stock and Yogo, 2005; Skeels and Windmeijer, 2018). Then we will corroborate the results of Skeels and Windmeijer (2018) for the more controversial case of two instruments, and finally, we will present results for the single instrument problem, for both linear and DCM. This extension is relevant for practice because, in many real situations, it is challenging even to get one proper instrument.

The rest of the chapter is organised as follows. Following this introduction, the methodological framework section is divided into two parts, beginning with an overview of the related literature, followed by a detailed account of the two state-of-the-art works for linear models (Stock and Yogo, 2005; Skeels and Windmeijer, 2018). Section 4.3 describes the alternative Monte Carlo method proposed here to test for weak instruments and its application to linear models, which serves as a validation, and the addition of the single instrument case. Section 4.4 presents the application of the proposed Monte Carlo method to test for weak instruments extended and adapted to DCM. Finally, section 4.5 presents our main conclusions.

4.2. Methodological Framework

In this section, we present an overview of the literature about the weak instruments problem, highlighting the main findings to date for linear models and DCM. Then, we describe the state of the art to detect weak instrument in the case of linear models.

4.2.1. *Literature overview*

The estimation of inconsistent parameters in endogenous models due to the weakness of the proposed instruments was a research gap pending in econometric modelling until the 1980s. Phillips (1989), seems to have been the first to put attention on the distributional consequences of using weak instruments. His research highlighted the severe problems that may arise when the instruments are not able to satisfy the relevance condition. If the instruments are not correlated "enough" with the endogenous variables, then the model is only partly identified and conventional asymptotic breaks down. Later, Nelson and Startz (1990a, 1990b) and Bound *et al*. (1995) warned about this severe econometric anomaly and the consequences of not correcting it. At the end of the 90s, Shea (1997) and Godfrey (1999), suggested using the coefficient of determination (R^2) of the first stage of the two-stage least-squares (TSLS) method, as a measure to establish the weakness/strength of instruments to correct for endogeneity. Initially, R^2 was considered a useful measure of the relevance condition for univariate models, with the warning that it could be misleading when there are multiple endogenous variables. Staiger and Stock (1997) were the first to define a "*rule of thumb*", from an asymptotic distribution, to test for weak instruments. This rule established that an instrument is weak if the first-stage F-statistic is less than ten. On the other hand, Zivot *et al*. (1998) recommended checking the performance of the first-stage regression and then making an inference based on the likelihood ratio (LR) or Lagrange multiplier (LM) statistics. To apply the tests appropriately, they suggested a correction based on the degrees-of-freedom of the model in the overidentified case.

More recently, Stock and Yogo (2005) formalized further the analysis of the weak instruments problem for linear regression. Their fundamental contribution was to determine the *critical values* (CV) for identifying weak instruments based on two unambiguous criteria: *relative bias* (RB) and *size distortion* (SD) of the Wald (1943) test. If there is one endogenous variable, then the CV are obtained using the first-stage F-statistic as a performance measure to test whether the instruments are weak (Sanderson and Windmeijer, 2016). For the case of two or more endogenous variables, the Cragg and Donald (1993) statistic is used. The CV tabulated by Stock and Yogo (2005) depend on the estimator (TSLS, LIML or Fuller-*k*) that the modeller is using, the number of instruments, the number of endogenous regressors, and how much bias or distortion (5%, 10% or more) the modeller considers tolerable. However, a practical difficulty of this approach is its analytic derivation, because the CV are reached from the evaluation of a non-straightforward integral, requiring Monte Carlo simulation to solve it.

To address this limitation, Skeels and Windmeijer (2018) have recently proposed an analytical closed-form solution of the integral, which they evaluated numerically using MATLAB (MathWorks, 2016). This finding allowed them to extend the CV results of Stock and Yogo (2005) to include more variation in the number of instruments and degree of RB. Andrews *et al*. (2019) did a complete literature review about the detection of weak instruments and the construction of robust confidence sets, focusing mainly on their practical importance.

Research about the identification of weak instruments in DCM is scarce. Dufour and Wilde (2018) used Monte Carlo experiments to measure the performance of the Wald and LR tests when Probit models are under the effects of weak instruments. Also, some findings have determined that the Wald test exhibits large levels of distortion (over-reject the null hypothesis or Error type I) under weak instruments in DCM, implying that this test is unreliable (Magnusson, 2007; Dufour and Wilde, 2018). As we show later, our approach to the problem is different. Following an idea preliminarily explored by Guevara and Navarro (2015), we reconstruct the empirical distribution of the F-statistics under weak instruments in DCM and are able to find CV depending on the number of instruments used and the level of RB the analyst is willing to tolerate.

4.2.2. *State of the art on testing for weak instrument in linear models*

Below, we describe the framework proposed for linear models by Stock and Yogo (2005), and improved by Skeels and Windmeijer (2018), in some depth. We will use their findings as a benchmark for validation. The methodological development is shown only for a single endogenous variable in a linear model, given that this is the case we can extend for DCM using the Monte Carlo approach described in Section 4.3.

Consider the linear model shown in (4.1) and the reduced form for the explanatory variable *x* in (4.2):

$$y = \theta x + \epsilon \qquad (4.1)$$

$$x = \pi z + \varDelta \qquad (4.2)$$

where y, $\varDelta$ and ϵ are $N*1$ vectors, N is the number of observations. x is a matrix of regressors of dimension $N*k_x$, where k_x corresponds to the number of regressors; on the other hand, z is an $N*k_z$ matrix of exogenous variables, hereafter labelled as instruments (or instrumental variables), and θ and π are parameters to estimate.

For this model, endogeneity arises when the conditional expectation $E(\epsilon|\varDelta) \neq 0$; therefore, x is correlated with ϵ in (1) because $\varDelta$ and ϵ have some level of correlation (ρ). Note also that z is correlated with x through (4.2), but not with ϵ in (4.1); thereby, z is a proper instrument for x. The value of the parameter π represents the level of weakness/strength of the instrument (Staiger and Stock 1997).

Stock and Yogo (2005) proposed two quantitative criteria to detect weak instruments. Both criteria were derived for the estimator of one or more endogenous regressors. The first is based on the maximum estimator bias, and was called *relative bias* (RB), whereas the second is related to the maximum Wald test's *size distortion* (SD). Here we focus on the former, which is defined as the absolute value of the ratio between the bias of the corrected model and the bias of the endogenous model, as shown in (4.3):

$$RB = \left|\frac{E[\hat{\theta}_{TSLS}]-\theta}{E[\hat{\theta}_{OLS}]-\theta}\right| \qquad (4.3)$$

where θ is the population (or true) parameter, $\hat{\theta}_{TSLS}$ is the parameter estimated using TSLS (i.e., corrected for endogeneity) and $\hat{\theta}_{OLS}$ is the endogenous (non-corrected) parameter determined using Ordinary Least Square (OLS). Stock and Yogo (2005) used weak instruments asymptotic to determine the degree of RB corresponding to a 5% significance level of the first-stage F-statistic, for the null hypothesis that the coefficients of all instruments were equal to zero.

The F-statistic, in general, is used to test the null hypothesis (H_0) that a linear restriction on the model parameters is true. The statistic is calculated from the *sum of squared residuals* of both the restricted model ($SSRR$) and the unrestricted model ($SSRU$), that is, if the restrictions are not imposed. The restricted model, in this case, corresponds to a model where the coefficients of all instruments are zero and, thus, the F-statistic is calculated as follows:

$$F = \frac{(SSRR - SSRU)}{SSRU} \frac{(N-K)}{k_z} \xrightarrow{d} \chi^2_{k_z} , \tag{4.4}$$

where K and k_z stand for the number of variables in the unrestricted model, and the number of restrictions imposed, respectively. The intuition behind the test is that if the restrictions are true, $SSRR$ should be similar to $SSRU$ and thus the statistic should be close to zero. On the contrary, the null hypothesis can be rejected (Ortúzar and Willumsen, 2011). The F-statistics follows a χ^2_{df} with degrees of freedom equal to the number of instruments k_z.

The critical values determined by this approach (Stock and Yogo, 2005; Skeels and Windmeijer, 2018) for a single endogenous regressor are shown in Table 4-1. We reproduce these values in detail because we will use them to contrast our results on weak instruments for DCM.

k_z	RB (Stock and Yogo, 2005)							RB (Skeels and Windmeijer, 2018)						
	0.01	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.01**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**
2	-	-	-	-	-	-	-	11.57	9.02	7.85	7.14	6.61	6.19	5.83
3	-	13.91	9.08	-	6.46	-	5.39	46.32	13.76	9.18	7.52	6.60	5.96	5.49
4	-	16.85	10.27	-	6.71	-	5.34	63.10	16.72	10.23	7.91	6.67	5.88	5.32
5	-	18.37	10.83	-	6.77	-	5.25	72.55	18.27	10.78	8.11	6.71	5.82	5.19
6	-	19.28	11.12	-	6.76	-	5.15	78.59	19.19	11.08	8.21	6.70	5.75	5.09
7	-	19.86	11.29	-	6.73	-	5.07	82.75	19.79	11.25	8.25	6.67	5.69	5.01
8	-	20.25	11.39	-	6.69	-	4.99	85.78	20.20	11.36	8.26	6.64	5.63	4.93
9	-	20.53	11.46	-	6.65	-	4.92	88.07	20.49	11.42	8.25	6.60	5.58	4.87
10	-	20.74	11.49	-	6.61	-	4.86	89.86	20.70	11.46	8.24	6.56	5.52	4.81
11	-	20.90	11.51	-	6.56	-	4.80	91.30	20.86	11.49	8.22	6.53	5.48	4.76
12	-	21.01	11.52	-	6.53	-	4.75	92.47	20.99	11.50	8.20	6.49	5.43	4.71
13	-	21.10	11.52	-	6.49	-	4.71	93.43	21.08	11.50	8.17	6.46	5.39	4.67
14	-	21.18	11.52	-	6.45	-	4.67	94.25	21.16	11.50	8.15	6.42	5.36	4.63
15	-	21.23	11.51	-	6.42	-	4.63	94.94	21.22	11.49	8.13	6.39	5.32	4.59
20	-	21.38	11.45	-	6.28	-	4.48	97.25	21.37	11.44	8.02	6.26	5.18	4.45
25	-	21.42	11.38	-	6.18	-	4.37	98.53	21.42	11.38	7.93	6.16	5.08	4.35
30	-	21.42	11.32	-	6.09	-	4.29	99.31	21.42	11.31	7.85	6.08	5.00	4.27

Table 4-1. *Critical values for a single endogenous regressor in linear models*

Various things can be noted in Table 4-1. The first is that the critical values grow with k_z and decrease with RB. This implies that as more instruments are used, a larger F-statistic of the first stage is needed to attain a certain degree of RB and, the more tolerant the researcher is with the RB, the less demanding the F-statistic becomes. Second, note that Stock and Yogo do not provide CV for RB of 0.01, 0.15 and 0.25. Third, the CV of Stock and Yogo (2005) start at k_z=3, whereas those of Skeels and Windmeijer (2018) start at k_z=2. Stock and Yogo (2005) claim that the need for $k_z \geq 3$ arises because the TSLS estimator does not have a finite-sample first moment when the number of instruments is one or two (i.e., the mean of the estimator could be infinite). Skeels and Windmeijer (2018) report results from k_z=2 with a caveat, motivating the latter value as an *ad-hoc* approximation supported by the arguments of Kinal (1980) regarding finite biases.

4.3. A Monte Carlo Method to Test for Weak Instruments in Linear Models

This section extends and improves a methodological approach proposed by Guevara and Navarro (2015). Instead of relying on asymptotic theory, we use Monte Carlo simulation to obtain an empirical distribution of the first stage's F-statistics of the control function (CF) method for the desired RB. We then retrieve the percentile 95th as the critical value. In the simulation, the desired RB is achieved by modifying one of the model parameters. We apply the correction using the CF approach instead of the TSLS[13].

To validate our Monte Carlo approach, we will contrast its results with those obtained from analytical methods (Table 4-1). Besides, and as explained above, we will consider the single instrument problem. For this, we will consider the linear model in (4.5) and the reduced form in (4.6) for an explanatory variable t, regarded as the only endogenous variable in this example. The aim is to emulate the analytic process of Stock and Yogo (2005), later improved by Skeels and Windmeijer (2018).

$$y = \theta_0 + \theta_c c + \theta_t t + \epsilon \qquad (4.5)$$

$$t = \pi_0 + \pi_{z1} z_1 + \pi_{z2} z_2 + \mathit{\Delta} \qquad (4.6)$$

where c and t are the independent variables of the linear model, y is the dependent variable and ϵ is the error term; on the other hand, the reduced form (4.6) for the explanatory variable t is explained by the intercept term (π_0), a couple of instruments z_1 and z_2, and $\mathit{\Delta}$ (error term). For illustrative purposes, we will consider only two instruments, because this is the minimum number reported in the RB estimates by Skeels and Windmeijer (2018). Nevertheless, the CF method requires at least one instrument for each endogenous variable to correct for endogeneity. As we show later, the findings achieved for this validation process were estimated up to fifteen instruments, which seems to be an adequate maximum in practical terms.

The simulated data were generated for the maximum level of correlation (ρ=1) between ϵ and $\mathit{\Delta}$; in this way, t and ϵ are correlated in (4.5) and, therefore, endogeneity arises. The variable y was constructed as a function of c, t and ϵ, with coefficients $\theta_0 = \theta_t = \theta_c = 2$. $\mathit{\Delta}$, c and ϵ were simulated using independent and identically (*iid*) Normal (0,2) draws, whereas z_1 and z_2 distributed Normal (0,1) and the intercept term $\pi_0 = 1$. The difference between the variances is to give less variability to the instruments than to the other variables.

[13] In linear models, the CF or TSLS methods can be used as they yield consistent parameters that are, numerically, very similar. In DCM, CF is the most appropriate method because it allows to make forecasts. As we aim to apply the CF approach to DCM, but rely on the linear models results, we used CF in both cases.

We did some simulations changing the variance for the normal distributions and the values of the parameters θ. We corroborated that these changes do not affect the results of our simulation.

The validation stage is based on the estimation of two linear models: an endogenous and a corrected one, as we need to estimate the RB shown in (4.3). We use the superscripts *END* and *CF* to denote the origin of the vector of parameter estimates. The endogenous model is estimated by OLS following the functional form (4.7), that is, without correcting for endogeneity:

$$y = \hat{\theta}_0^{END} + \hat{\theta}_c^{END} c + \hat{\theta}_t^{END} t + \epsilon \quad (4.7)$$

To obtain the estimated parameters of the corrected model, we use the two-stage CF approach (Wooldridge, 2010). In the first stage, shown in (4.8), the residuals ($\hat{\varphi}$) are obtained from the OLS regression of the endogenous t on the exogenous variables (c) and the instruments (z_1 and z_2). In the second stage, shown in (4.9), we estimate the linear model considering the residuals ($\hat{\varphi}$) as further explanatory variables:

$$t = \phi_0 + \hat{\phi}_{z1} z_1 + \hat{\phi}_{z2} z_2 + \hat{\phi}_c c + \varphi \xrightarrow{OLS} \hat{\varphi} = t - \hat{t} \quad (4.8)$$

$$y = \hat{\theta}_0^{CF} + \hat{\theta}_c^{CF} c + \hat{\theta}_t^{CF} t + \hat{\theta}_{\hat{\varphi}}^{CF} \hat{\varphi} + \epsilon \quad (4.9)$$

The estimation of the parameters of both models requires an iterative process, modifying the power of the instruments by adjusting π_{z1} and π_{z2} in (4.6) until reaching the desired RB (RB^{obj}). A flowchart of this process is shown in Figure 4-1. The subscript j is used to highlight that π_z in (4.6) is changed j times until it reaches the desired $\overline{RB}^j$, which corresponds to the average of the RB_m^j.

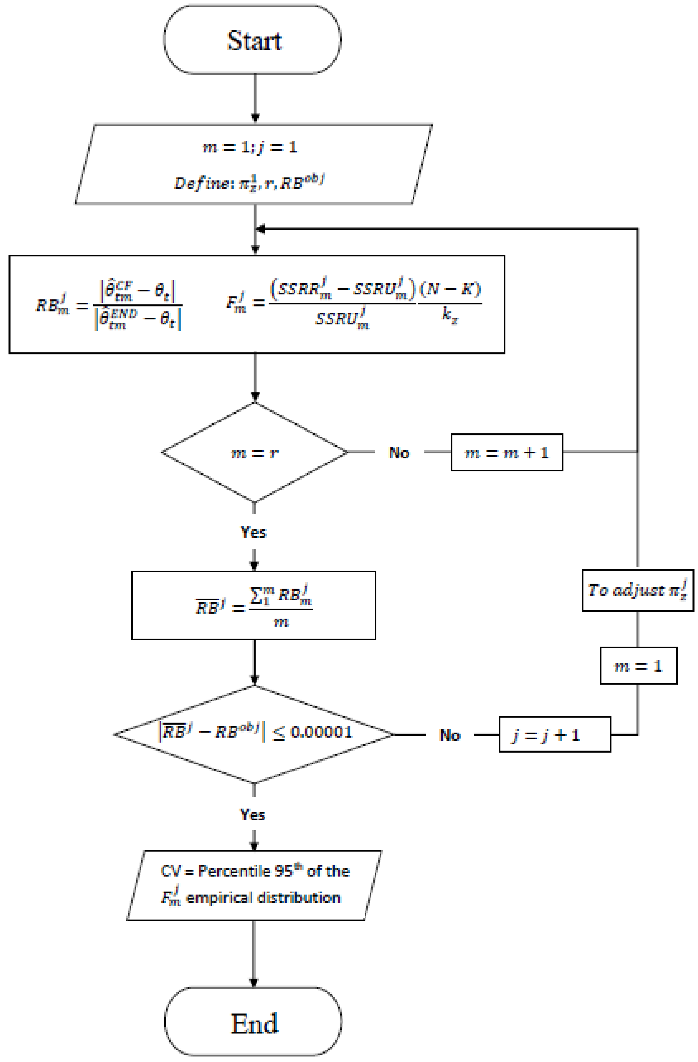

Figure 4-1. *Iterative process flowchart to reach the* $\overline{RB}^j$ *desired in linear models*

For illustrative purposes, Figure 4-2 shows the empirical distribution of the first-stage F-statistic for the CF approach in our Monte Carlo simulation, for each of the *m* replications in the linear models. The abscissa corresponds to the estimation of the first-stage F-statistic for the CF approach, for each repetition. The ordinate indicates the F-statistic density for all the replications. The initial value of π_z, RB^{obj} and the maximum number of times that the process will be repeated (r), must be defined by the modeller. The subscript *m* indicates the number of the repetition. The process is repeated until *m* is equal to *r* for a set of fixed instrument vectors and a given value of π_z^j, which was the same for all instruments analysed. For each π_z^j, the RB_m^j and F_m^j statistic are estimated *m* times (i.e., as many times as the experiment is repeated). The approach allows obtaining an empirical distribution of F_m^j, from the *m* repetitions, enabling us to compute the percentile 95th. This value corresponds to the critical value for the $\overline{RB}^j$ that the modeller is willing to tolerate for the respective number of instruments k_z used to correct endogeneity. Note that the definition of RB_m^j in Figure 4-1 is given by (4.3).

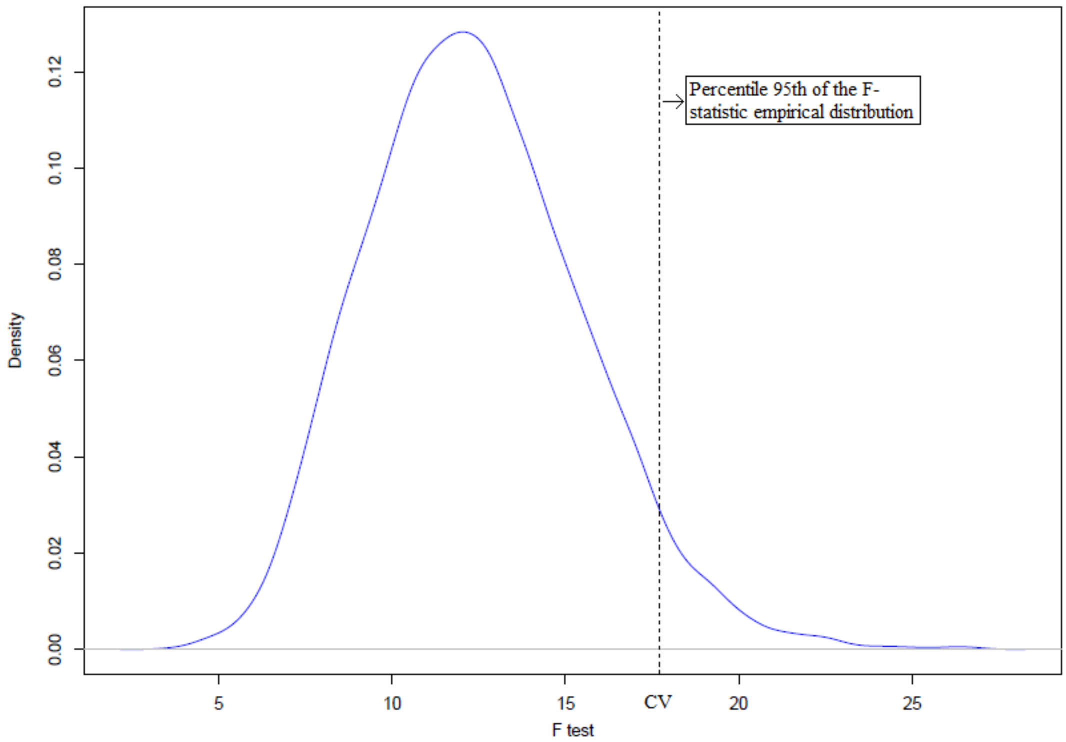

Figure 4-2. *Empirical distribution of the F-statistic in linear models*

The F-statistic empirical distribution is drawn as a blue line in Figure 4-2. The critical value is represented as a black vertical dashed line and corresponds to the percentile 95th of the F-statistic empirical distribution; this is the value proposed for the desired RB. The 95th percentile is preferred to any other statistical measure, such as the mean or the median, because it is the boundary value defining the zone of acceptance/rejection of the null hypothesis for a nominal level of 5% of the F-statistic empirical distribution.

We used the iterative process represented in Figure 4-1 for estimating the CV, and then compared our values with those of Stock and Yogo (2005), and Skeels and Windmeijer (2018), for linear models using analytical methods. Our critical values were reached for a sample size of 10.000 individuals. The number of repetitions can vary depending on the precision wanted and the computational time one is willing to invest. For this analysis we considered a precision of at least 0.05, enabling us to report until one decimal point of the CV estimates. This was achieved by means of 78,000 simulations, which were subsampled in sets of 100 repetitions from which we estimated the mean, median and 95% confidence interval (CI) for the CV, as summarized in Table 4-2. The estimates show, first, that the means and medians in linear models are almost identical (as expected), and that the 95% CI have widths that vary between 0.9 and 3.72 points. Comparing these simulated results with the values reported in Table 4-1, we note that the mean and median in Table 4-2 are virtually the same as the values in Table 4-1.

Number of instruments	RB = 0.05		RB = 0.10		RB = 0.15		RB = 0.20		RB = 0.25		RB = 0.30	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
	[CI]		[CI]		[CI]		[CI]		[CI]		[CI]	
2	8.9	8.9	7.6	7.6	6.9	6.8	6.5	6.4	6.1	6.0	5.8	5.8
	[7.37 - 10.82]		[5.89 - 9.41]		[5.56 - 8.57]		[5.2 - 8.12]		[4.88 - 7.52]		[4.47 - 7.24]	
3	13.5	13.5	9.0	8.9	7.4	7.4	6.5	6.5	5.9	5.8	5.4	5.4
	[11.8 - 15.52]		[7.77 - 10.51]		[6.24 - 8.76]		[5.43 - 7.76]		[4.87 - 7.07]		[4.44 - 6.57]	
4	16.4	16.3	10.0	9.9	7.7	7.7	6.5	6.5	5.7	5.7	5.2	5.2
	[14.73 - 18.24]		[8.8 - 11.14]		[6.65 - 8.77]		[5.61 - 7.54]		[4.87 - 6.7]		[4.42 - 6.09]	
5	18.0	18.0	10.6	10.5	8.0	7.9	6.6	6.6	5.7	5.6	5.1	5.1
	[16.52 - 19.89]		[9.54 - 11.78]		[7.1 - 8.99]		[5.81 - 7.56]		[4.96 - 6.54]		[4.41 - 5.86]	
6	18.9	18.9	10.8	10.8	8.0	8.0	6.6	6.5	5.6	5.6	5.0	4.9
	[17.51 - 20.57]		[9.91 - 11.97]		[7.26 - 9.09]		[5.9 - 7.46]		[5.01 - 6.41]		[4.44 - 5.75]	
7	19.6	19.6	11.1	11.0	8.1	8.1	6.6	6.6	5.6	5.6	4.9	4.9
	[18.19 - 21.11]		[10.08 - 12.17]		[7.4 - 9.07]		[5.9 - 7.41]		[4.97 - 6.37]		[4.34 - 5.65]	
8	20.0	20.0	11.2	11.2	8.2	8.1	6.6	6.6	5.5	5.5	4.9	4.8
	[18.73 - 21.56]		[10.31 - 12.23]		[7.43 - 9.02]		[5.89 - 7.28]		[4.94 - 6.2]		[4.3 - 5.49]	
9	20.4	20.4	11.3	11.3	8.2	8.1	6.5	6.5	5.5	5.5	4.8	4.8
	[19.2 - 21.77]		[10.41 - 12.18]		[7.39 - 8.96]		[5.8 - 7.27]		[4.84 - 6.11]		[4.23 - 5.38]	
10	20.6	20.6	11.3	11.3	8.1	8.1	6.5	6.5	5.4	5.4	4.7	4.7
	[19.39 - 21.89]		[10.5 - 12.18]		[7.43 - 8.87]		[5.86 - 7.14]		[4.91 - 6.01]		[4.23 - 5.28]	
11	20.9	20.9	11.4	11.3	8.1	8.1	6.5	6.5	5.4	5.4	4.7	4.7
	[19.69 - 22.12]		[10.53 - 12.19]		[7.44 - 8.84]		[5.89 - 7.11]		[4.87 - 6.03]		[4.18 - 5.28]	
12	21.1	21.1	11.4	11.4	8.1	8.1	6.4	6.4	5.4	5.3	4.7	4.6
	[19.98 - 22.3]		[10.63 - 12.23]		[7.46 - 8.8]		[5.89 - 7.04]		[4.89 - 5.99]		[4.19 - 5.21]	
13	21.2	21.2	11.4	11.4	8.1	8.1	6.4	6.4	5.3	5.3	4.6	4.6
	[20.15 - 22.39]		[10.7 - 12.25]		[7.49 - 8.79]		[5.88 - 6.98]		[4.86 - 5.88]		[4.18 - 5.11]	
14	21.4	21.3	11.4	11.4	8.1	8.1	6.4	6.4	5.3	5.3	4.6	4.6
	[20.27 - 22.47]		[10.72 - 12.25]		[7.46 - 8.79]		[5.85 - 6.97]		[4.82 - 5.78]		[4.1 - 5.06]	
15	21.4	21.4	11.4	11.4	8.1	8.1	6.3	6.3	5.3	5.3	4.5	4.5
	[20.39 - 22.57]		[10.74 - 12.24]		[7.48 - 8.73]		[5.88 - 6.88]		[4.84 - 5.74]		[4.14 - 4.99]	

Table 4-2*. Mean, median and CI for the CV of first-stage 95% F-statistic to detect weak instruments for a single endogenous regressor in linear models with Monte Carlo method*

Formally, then, we cannot reject the null hypothesis that the values obtained with our approach are equal to those reported by Stock and Yogo (2005) and Skeels and Windmeijer (2018). This serves as a validation of the proposed Monte Carlo approach to the problem.

To further illustrate the case, Figure 4-3 shows boxplots[14] for CV and confidence intervals as a function of the number of instruments, for RB = 0.05, together with the values reported by Skeels and Windmeijer (2018). The black and blue dots depict the mean and median (50th percentile) of all repetitions of the experiment in each boxplot. The abscissa shows the number of instruments k_z, and the ordinate corresponds to the CV obtained in the validation process. The minor differences shown can be attributed to sampling and simulation errors, and the difficulty involved in correcting econometric models with weak instruments. Therefore, the hypothesis that the CV coming from our simulation approach are the same as those proposed by Skeels and Windmeijer (2018) for linear models, cannot be rejected. This confirms the validity of the proposed Monte Carlo approach in the case of linear models.

[14] A boxplot is a standardized way of displaying empirical sampling distributions obtained from simulation and consists of a rectangle with bottom and topside at the levels of the 1st and 3rd quartile (i.e., 25th and 75th percentiles). The distance between the 1st and 3rd quartile is known as interquartile range (Tukey, 1977).

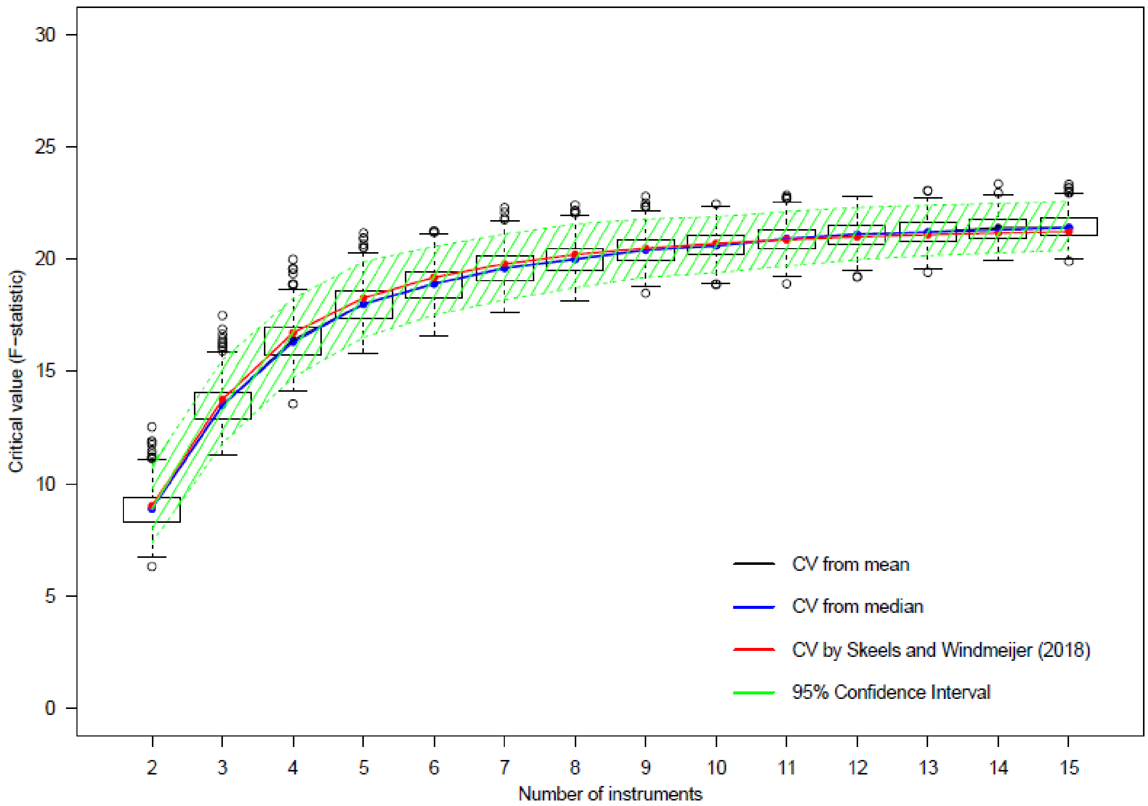

Figure 4-3. *Boxplot for CV and CI by number of instruments and RB 0.05 for linear models*

As discussed before, Skeels and Windmeijer (2018) were able to determine the CV for k_z=2 using the RB criteria, extending the work of Stock and Yogo (2005). This finding was significant, but it has not escaped controversy and it is even stated by the authors as an *ad-hoc* approximation. The problem is that the asymptotic analysis performed by these authors preclude determining the moments for the case of two instruments (Stock and Yogo, 2005; Angrist and Pischke, 2009). Notwithstanding, our results with the Monte Carlo approach are fully consistent with those attained by Skeels and Windmeijer (2018) using asymptotic theory.

Beyond the k_z=2 case, a more relevant situation for practice would be to provide recommendations for the single instrument case (namely k_z=1), since in most practical applications finding even a single instrument is extremely hard. In this sense, it would be interesting to know if the *rule of thumb* (F-statistic is less than ten), or if values “projected” from the results available for more instruments, would work. Looking at Tables 4-1, 4-2 and Figure 4-3, one would be tempted to say that this is the case, but a formal demonstration is missing.

However, if k_z=2 was controversial, k_z=1 seems out of the question. In their Appendix D, Skeels and Windmeijer (2018) explored the function's performance that defined their approximation for the case k_z=1 finding CV that were significantly larger than those suggested by a simple extrapolation of the values in Table 4-1. The CV for k_z=1 reported by Skeels and Windmeijer (2018) come from a Monte Carlo analysis for RB which values are 0.01, 0.05, 0.10 and 0.20.

As our Monte Carlo method has no limits regarding the number of instruments, we can analyse the case of k_z=1 and a single endogenous variable. The mean, median and confidence intervals from our simulation for the critical values, in this case, are reported in Table 4-3. Besides, the boxplots displayed in Figure 4-4 are drawn for the simulation conditions described above for k_z=1.

Number of instruments	RB = 0.05	RB = 0.10	RB = 0.15	RB = 0.20	RB = 0.20	RB = 0.30
	Mean Median [CI]	Mean Median [CI]	Mean Median [CI]	Mean Median [CI]	Mean Median [CI]	Mean Median [CI]
1	40.7 41.1 [28.26 – 46.97]	25.3 27.1 [15.44 – 32.84]	20.2 19.9 [9.35 – 29.63]	18.9 18.7 [6.88 – 29.20]	17.9 18.5 [6.48 – 24.40]	16.6 17.1 [6.48 – 21.57]

Table 4-3. *Mean, median and CI for the CV of first-stage 95% F-statistic to detect weak instruments for a single endogenous regressor in linear models for k_z=1*

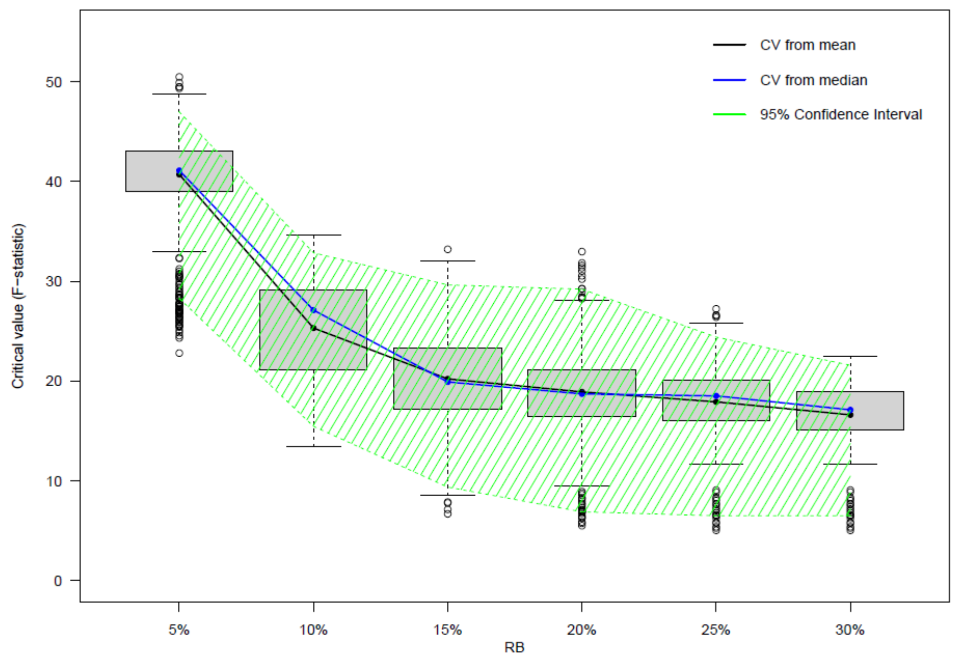

Figure 4-4. *Boxplot for CV when k_z =1 and RB=0.05 in linear models*

The critical values reached with our empirical approach for k_z=1 are far from those that would be suggested by a simple extrapolation of the values in Table 4-1. However, the CV in Table 4-3 are close to those reported in Table A1 Appendix D of Skeels and Windmeijer (2018) when k_z=1. These results imply that the "*rule of thumb*" that establish that an instrument is weak if the first-stage F-statistic is less than ten (Staiger and Stock, 1997) is far from being enough at least for this case.

We also investigated the effects of weak instruments on the estimators. We did simulations for the case of RB=0.05 and k_z=5. We appeal to the iterative process in Figure 4-1 to adjust π_z until reaching the mean and median reported in Table 4-2 (18.0 and 18.0, respectively) instead of the desired RB. We did more simulations for π_z values higher and smaller than that allowed reaching the mean and median reported in Table 4-2. In this way, we obtained empirical estimator distributions with and without weak instruments, as shown in Figure 4-5, where the abscissa corresponds to the $\hat{\beta}_t^{CF}$, the real value of which is 2.0 (β_t), and it is shown as a black dashed line. On the other hand, the ordinate indicates the estimator density (also known as estimator distribution). The red line corresponds to the estimator's sample distribution for the critical value when k_z=5 and RB=0.05. Under these conditions, the estimator's sample distribution still retains the "*bell*" shape of Normal distribution and it does not deform. Rather, by decreasing the F-statistic, the variance increases slightly; however, the correction is not better.

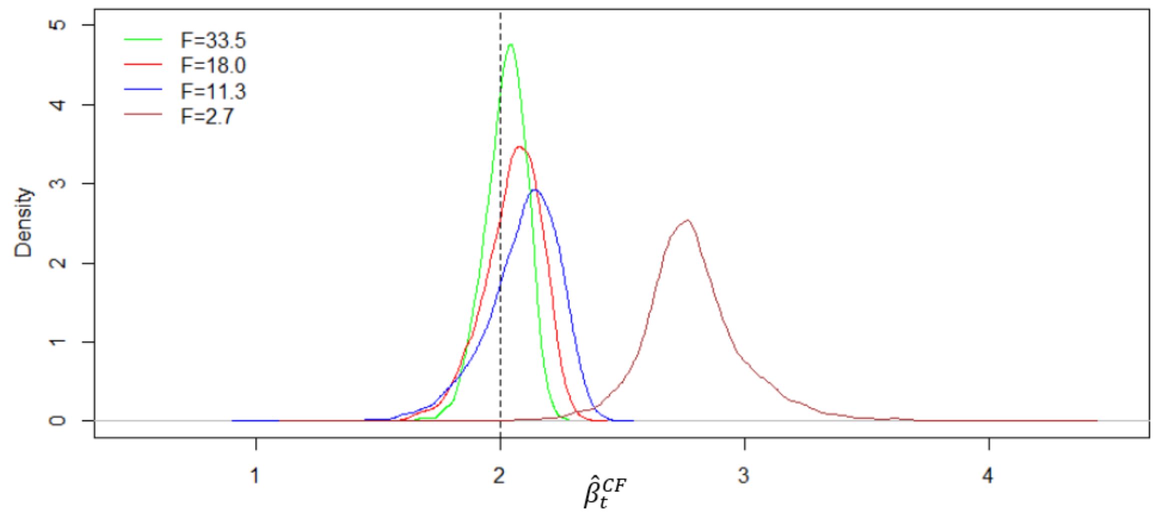

Figure 4-5. *Effect of the weak instruments on the estimator distribution in linear models*

4.4. Application of Monte Carlo Method to Test for Weak Instruments in DCM

To extend and adapt our empirical methodology to the case of DCM, we started with a simple binary choice model. We assumed that the utility function U_{in} for alternative *i* and individual *n* followed the functional form shown in (4.10).

$$U_{in} = ASC_i + \beta_c c_{in} + \beta_t t_{in} + \varepsilon_{in} \quad (4.10)$$

where, c_{in} and t_{in} are explanatory variables of the model, ε_{in} is a Normal distributed error term, and ASC_i is the alternative specific constant for alternative *i*. For simulation purposes, without loss of generality we assumed that the values of the parameters in (10) were as follows: $ASC_i = 2$, $\beta_t = 5$ and $\beta_c = 2$. As the error term ε_{in} distributes Normal, the choice model is formally a Probit model (Train, 2009). Notwithstanding, to simplify estimation we estimated it as a binary Logit. Note that Lee (1982), Ruud (1983) and Cramer (2007) show that this is a minor issue, as it does not compromise the possibility to obtain consistent parameters up to a scale.

The data were generated using Monte Carlo simulation for a discrete choice model that suffers from endogeneity. Although our analysis was done for a binary case, as we show below, it can easily be extended to multinomial DCM. We built our simulation experiment for a single endogenous variable; in particular, we assumed that t_{in} in (4.10) was endogenous. Thus, t_{in} was constructed as a function of an intercept term (α_0), an instrument's matrix (z_{in}) and an error term (ξ_{in}), as shown in (4.11). Again, we considered two instruments (z_{1in} and z_{2in}). For the case of DCM, the findings achieved were estimated up to fifteen instruments, which seems to be an adequate maximum in practical terms.

$$t_{in} = \alpha_0 + \alpha_{z1} z_{1in} + \alpha_{z2} z_{2in} + \xi_{in} \quad (4.11)$$

As we generated correlation between the terms ε_{in} and ξ_{in}, $[corr(\varepsilon_{in}, \xi_{in}) = \rho \neq 0]$, t_{in} is correlated with ε_{in} in the utility function (4.10) and, by definition, endogeneity arises. We simulated the data for ρ=1, that is, the maximum level of correlation. c_{in}, ξ_{in} and ε_{in} were simulated using independent and identically (*iid*) Normal (0,2) draws, whereas z_1 and z_2 distributed Normal (0,1) and the intercept term was set as $\alpha_0 = 1$.

Following the same methodology used in the case of linear models above, we estimated two models: one endogenous and one corrected. Again, the superscripts *END* and *CF* are used for the parameters coming from the endogenous and corrected models, respectively. The former is estimated using the functional form (4.12), which is not corrected for endogeneity.

$$U_{in} = \widehat{ASC}_i^{END} + \hat{\beta}_c^{END} c_{in} + \hat{\beta}_t^{END} t_{in} + \varepsilon_{in} \qquad (4.12)$$

To correct for endogeneity, we used the two-stage CF method (Wooldridge, 2010). If the simultaneous procedure (also called Maximum Likelihood) was preferred, the reader can refer to Train (2009). The two-stage approach was used here because it involves a lower computational cost than the simultaneous procedure.

The first stage, shown in (4.13) consists in obtaining the residuals ($\hat{\delta}_{in}$) from the OLS regression of t_{in} on the exogenous variables (c_{in}) and the instruments (z_{in}). In the second stage, the DCM is estimated considering $\hat{\delta}_{in}$ (from the first stage) as an extra explanatory variable in the utility function (4.14). This two-stage procedure does not compromise estimator's consistency, which is guaranteed by the Slutsky theorem (Ben-Akiva and Lerman, 1985).

$$t_{in} = \hat{\gamma}_t + \hat{\gamma}_{z1} z_{1in} + \hat{\gamma}_{z2} z_{2in} + \hat{\gamma}_c c_{in} + \delta_{in} \xrightarrow{OLS} \hat{\delta}_{in} = t_{in} - \hat{t}_{in} \qquad (4.13)$$

$$U_{in} = \widehat{ASC}_i + \hat{\beta}_c^{CF} c_{in} + \hat{\beta}_t^{CF} t_{in} + \beta_{\hat{\delta}} \hat{\delta}_{in} + \tilde{\varepsilon}_{in} \qquad (4.14)$$

The iterative process flowchart to determine critical values in this case is shown in Figure 4-6 and it is the same than the case of linear models. However, as correcting for endogeneity in DCM implies a change of scale in the estimators (Guevara and Ben-Akiva, 2012), it is more convenient to check the ratio among parameters (i.e., β_t/β_c) rather than the parameters directly when making comparisons. Therefore, the equation to calculate the RB_m^j for DCM in Figure 4-6 is different from the one shown in Figure 4-1.

We continued considering that a precision of 0.05 or less for the CV was enough for our DCM results. This allows us to report until one decimal for the CV estimates. We also kept the same simulation settings designed for the case of linear models regarding sample size and number of repetitions. In this way, the estimates from the mean, median and 95% confidence interval (CI) for the CV, for a single endogenous regressor in DCM, are summarized in Table 4-4.

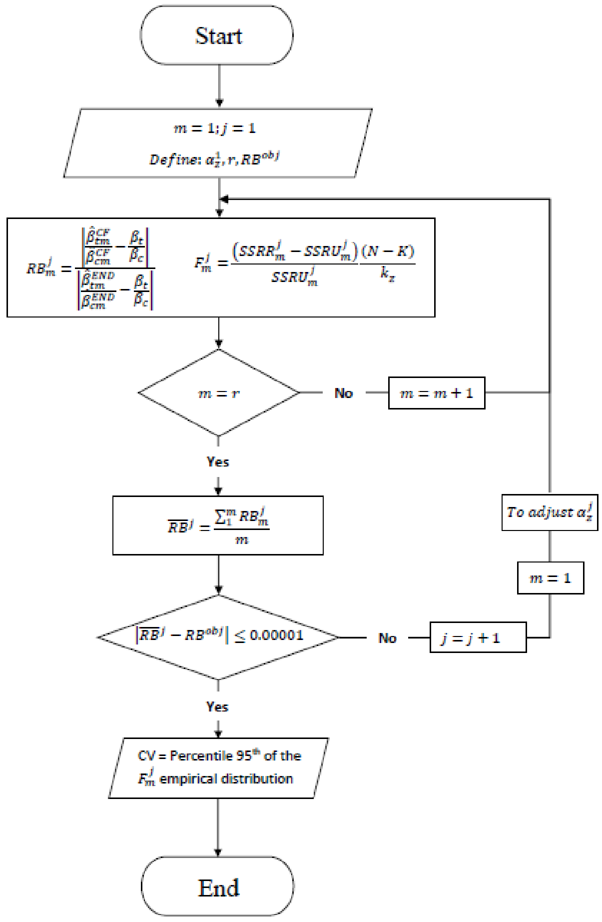

Figure 4-6. *Iterative process flowchart to reach the $\overline{RB}^j$ desired in DCM*

As can be seen, the median and mean from the simulation tends to be very similar again. Notwithstanding, we recommend using the CV associated with the median, as this does not consider the exact locations of the values in a dataset, only their relative standing when they are ordered. Therefore, the median allows getting a better idea of a "typical" value, given that it is not skewed by extreme observations or outliers (Washington *et al*. 2020). The CI were estimated using the percentile empirical distribution directly (in our case 2.5% and 97.5%), to represent the confidence interval (CI) at the 5% significance level (Davison and Hinkley 1997). In this way, if the F-statistic is lower than the critical value in Table 4-4, then the instruments used to correct for endogeneity in DCM can be considered weak. Our approach allows determining the CV for DCM in the case of k_z=2 to k_z=15 and RB from 0.05 to 0.30. The CV reported in Table 4-4 show, as expected, that as one moves across columns from left to right for each row, the CV becomes smaller. This finding is in line with that reported for linear models.

Figure 4-7 shows the boxplots for the CV estimated from k_z=2 to k_z=15 and RB=0.05. Again, the black and blue dots depict the mean and median (50th percentile) for all repetitions of the experiment.

Number of instruments	RB = 0.05		RB = 0.10		RB = 0.15		RB = 0.20		RB = 0.25		RB = 0.30	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
	[CI]		[CI]		[CI]		[CI]		[CI]		[CI]	
2	9.4	9.3	8.3	8.2	7.4	7.4	6.9	6.8	6.4	6.2	5.9	5.8
	[7.26 - 11.47]		[6.15 - 10.09]		[5.42 - 9.26]		[4.90 - 9.12]		[4.77 - 7.79]		[3.71 - 7.61]	
3	13.5	13.4	8.8	8.8	7.3	7.2	6.6	6.5	5.9	5.8	5.5	5.3
	[10.71 - 16.75]		[7.21 - 10.56]		[5.95 - 9.16]		[5.25 - 8.41]		[4.56 - 7.63]		[4.35 - 6.93]	
4	16.5	16.5	9.7	9.6	7.5	7.5	6.5	6.4	5.9	5.7	5.2	5.2
	[13.65 - 19.37]		[7.98 - 11.72]		[6.07 - 9.24]		[5.22 - 8.03]		[4.67 - 7.18]		[4.17 - 6.43]	
5	18.0	17.9	10.6	10.5	7.9	7.8	6.6	6.5	5.7	5.7	5.1	5.1
	[14.00 - 21.57]		[8.51 - 12.87]		[6.49 - 9.93]		[5.50 - 8.30]		[4.73 - 7.14]		[4.19 - 6.30]	
6	19.3	19.0	11.0	10.9	8.1	8.0	6.7	6.6	5.7	5.7	5.1	5.1
	[15.19 - 24.58]		[8.87 - 13.17]		[6.57 - 9.76]		[5.47 - 8.19]		[4.71 - 7.04]		[4.25 - 6.14]	
7	20.0	20.0	11.4	11.2	8.2	8.1	6.7	6.6	5.7	5.7	5.0	5.0
	[17.33 - 23.44]		[9.76 - 13.54]		[6.85 - 9.62]		[5.51 - 7.97]		[4.70 - 6.82]		[4.17 - 5.98]	
8	20.3	20.3	11.5	11.3	8.2	8.1	6.6	6.6	5.6	5.6	4.9	4.9
	[18.13 - 22.56]		[10.01 - 13.48]		[6.94 - 9.55]		[5.46 - 7.76]		[4.74 - 6.61]		[4.18 - 5.82]	
9	20.7	20.5	11.4	11.3	8.3	8.2	6.6	6.6	5.5	5.5	4.8	4.8
	[18.32 - 23.20]		[10.09 - 12.99]		[7.07 - 9.56]		[5.57 - 7.62]		[4.71 - 6.41]		[4.11 - 5.63]	
10	21.1	21.2	11.7	11.7	8.2	8.2	6.6	6.6	5.5	5.5	4.8	4.8
	[18.56 - 23.54]		[10.55 - 13.05]		[7.31 - 9.33]		[5.70 - 7.48]		[4.77 - 6.37]		[4.12 - 5.46]	
11	21.5	21.3	11.7	11.7	8.2	8.2	6.5	6.5	5.4	5.4	4.7	4.7
	[19.12 - 24.36]		[10.32 - 13.39]		[7.31 - 9.4]		[5.70 - 7.38]		[4.69 - 6.22]		[4.04 - 5.41]	
12	21.7	21.8	11.7	11.8	8.2	8.2	6.5	6.5	5.4	5.4	4.7	4.7
	[19.63 - 24.07]		[10.40 - 12.78]		[7.26 - 9.17]		[5.77 - 7.20]		[4.83 - 6.13]		[4.13 - 5.29]	
13	21.5	21.7	11.9	11.9	8.3	8.3	6.5	6.5	5.4	5.4	4.7	4.6
	[19.22 - 23.61]		[10.78 - 12.95]		[7.51 - 9.19]		[5.90 - 7.29]		[4.91 - 6.06]		[4.14 - 5.26]	
14	21.5	21.6	11.7	11.7	8.2	8.2	6.5	6.5	5.4	5.4	4.6	4.7
	[19.62 - 23.47]		[10.77 - 12.70]		[7.27 - 9.02]		[5.82 - 7.19]		[4.72 - 6.09]		[4.05 - 5.24]	
15	21.5	21.4	11.6	11.6	8.1	8.1	6.4	6.4	5.4	5.3	4.6	4.6
	[20.04 - 23.07]		[10.72 - 12.50]		[7.30 - 8.93]		[5.74 - 7.02]		[4.79 - 5.93]		[4.07 - 5.11]	

Table 4-4. *Mean, median and CI for the CV of first-stage 95% F-statistic to detect weak instruments for a single endogenous regressor in DCM with Monte Carlo method*

The CV for linear models estimated by Skeels and Windmeijer (2018) is shown as a red dot. The green dashed lines represent the upper and lower bounds of the CI. As can be seen, the mean and median reached in the simulation tend to be equal and differ in very few cases. Note also that the dots representing the mean and median (black and blue lines, respectively) are contained within the CI. Thereby, we cannot reject the null hypothesis that the mean and median are statistically equal to the CV reached by Skeels and Windmeijer (2018) for linear models. The figure also shows the presence of outliers in the empirical distribution of the CV. This can be, at least partly, attributed to the random process inherent to the Monte Carlo simulation. Besides, we consider that these outliers may be caused by the poor/slight correlation between the endogenous variable and the instruments in the CF approach's first stage. Additionally, this undesirable situation may lead to inconsistent estimates (Stock and Yogo, 2005), especially in DCM. As can be seen, many of these outliers occur when k_z=6 and k_z=7 in Figure 4-7. In the presence of outliers, we recommend using the median instead of the mean, because it allows getting a better idea of a "typical" value, as it is not skewed by extreme observations or outliers (Washington *et al*. 2020).

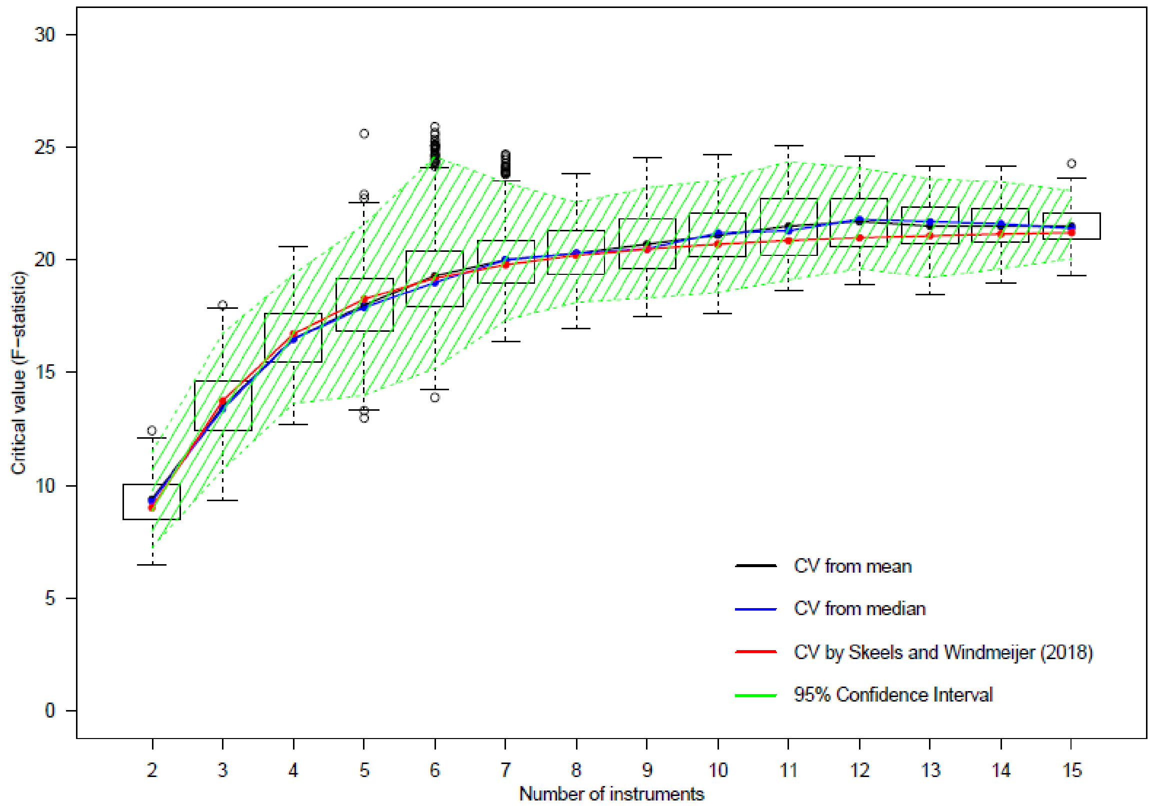

Figure 4-7. *Boxplot for CV and CI by number of instruments and RB=0.05 for DCM*

As we have commented already, in practice it is difficult to find proper instruments to correct for endogeneity; in fact, finding at least one of them may be difficult or even impossible (Guevara, 2015). So, in practical terms it would be useful to know the CV for k_z =1. Given that we applied our methodological approach for linear models and reached CV for k_z =1 and a single endogenous variable, we decided to test the same simulation conditions but applied to DCM. There are no reasons to think that our methodological approach would not work under these conditions, because it is well known that the CF method requires at least one instrument for each endogenous variable. The mean, median and CI for critical values from our empirical approach in the case of a single endogenous regressor in DCM are reported in Table 4-5 and shown in Figure 4-8.

Interestingly, the CV estimated for the DCM have the same order of magnitude of those achieved for linear models. Skeels and Windmeijer (2018) recommend that their findings for k_z =1 were useful for absolute values of RB less than 0.20, because the TSLS estimator in their simulation had a large standard deviation. But our method allows us to reach CV for RB up to 0.30. In any case, our results for k_z =1 in DCM are, in general, in line with those of Table A1 of Appendix D in Skeels and Windmeijer (2018).

Number of instruments	RB = 5%		RB = 10%		RB = 15%		RB = 20%		RB = 25%		RB = 30%	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
	[CI]		[CI]		[CI]		[CI]		[CI]		[CI]	
1	42.7	42.7	28.4	28.6	24.0	24.4	20.4	20.6	19.7	19.1	14.5	14.8
	[35.15 – 50.00]		[23.35 – 33.34]		[19.07 – 28.80]		[13.05 – 24.62]		[15.91 – 25.65]		[7.21 – 22.80]	

Table 4-5. *Mean, median and CI for the CV of first-stage 95% F-statistic to detect weak instruments for a single endogenous regressor in DCM with Monte Carlo Method for* k_z*=1*

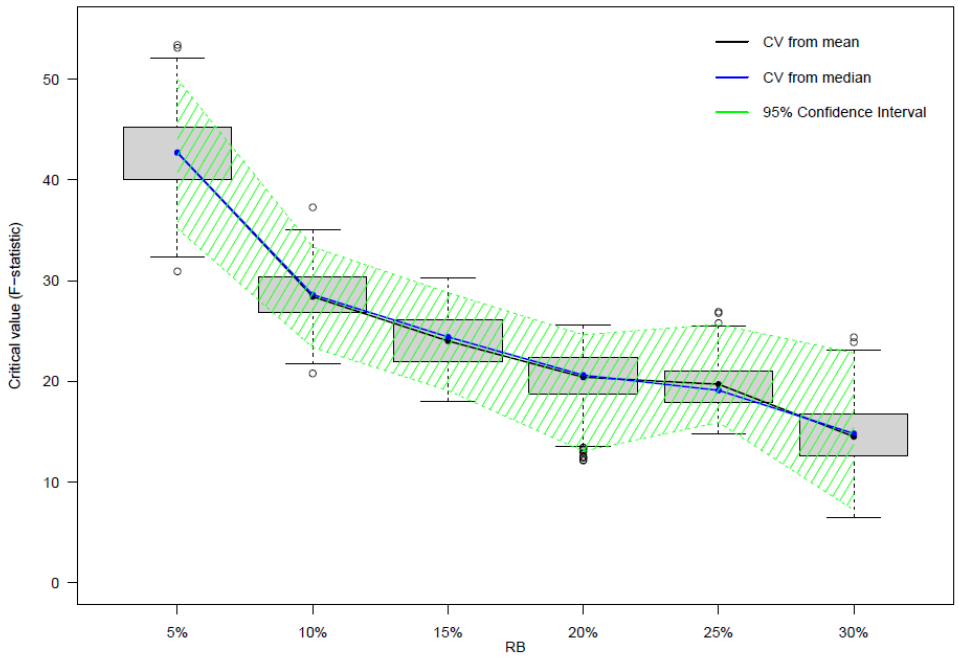

Figure 4-8. *Boxplot for CV when kz =1 in DCM*

There is no reason for not extending our methodological approach to the multinomial case. For illustrative purposes, we explored the impact of extending the analysis to five alternatives. We kept the same simulation conditions used above and analysed the case k_z=6 and RB=0.10[15]. From Table 4-4, we know that the mean of the binary case is 11.0 and the median is 10.9 for this case; the estimates of the mean and median for a 5 alternatives' case, k_z=6 and RB=0.10, are 11.04 and 10.93, respectively. Applying the t^*-statistics (Ortúzar and Willumsen, 2011, pages 341-342) to compare if the mean and median values from the binary and 5-alternative models are statistically the same, we found that H_0 can be accepted. Therefore, we conclude that our findings can indeed be extended to the case of logit models with multiple alternatives.

On the other hand, Figure 4-9 shows the effects of weak instruments on the estimators coming from DCM. The abscissa corresponds to the ratio $\hat{\beta}_t^{CF}/\hat{\beta}_c^{CF}$, the real value of which is β_t/β_c=2.5 and is shown as a black dashed line. The ordinate indicates the estimator density (distribution). Continuing with that done for linear models, we appeal to the iterative process in Figure 4-6 seeking to adjust α_z until reaching the mean and median reported in Table 4-4 (18.0 and 17.9, respectively) instead of the desired RB. As can be seen, when the instruments are strong (green line), the ratio of the estimator's sample distribution is centred on the real value (2.5) and has relatively symmetrical tails. However, as α_z decreases (and, therefore, the F-statistic), the correction's quality worsens, the variance increases, and the estimator's sample distribution deforms. This is a relevant finding because this behaviour does not happen in the case of linear models. The differences between linear models and DCM could be due to the scale's change in the estimators involved in the case of the DCM estimators (Guevara and Ben-Akiva, 2012).

[15] We only show the estimates for k_z=6 and RB=0.10 due to the computational cost involved in doing the exercise for all combinations of k_z and RB.

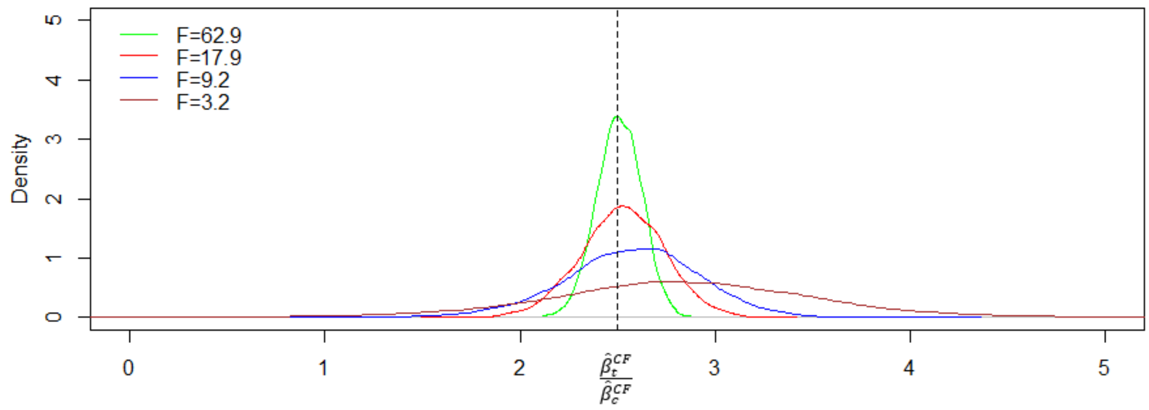

Figure 4-9. *Effect of the weak instruments on the estimator distribution in DCM*

Finally, we examined the asymptotic rejection rate of the weak instrument test, that is, the empirical *power* of the test as a function of the *effect size*. Power and effect size are the usual tools for the assessment of statistical analysis (Cohen, 1988). In our case, the power was calculated as the rejection rate of the Walt test (Wald, 1943) for a nominal level of 5%, whereas the effect size corresponds to the gradual increase in the weakness of the instruments. This way, if the effect size increases, it is expected that the rejection rate will also increase. Given that the two-stage estimation with the CF approach involves inconsistent estimates of the standard errors[16], we used the *willingness-to-pay space* approach (Train and Weeks, 2005). This procedure allows estimating the correct standard errors for the endogenous variables and to apply the Wald test correctly.

The bounds on the asymptotic rejection rate are plotted in Figure 4-10 for RB = 0.10 and 0.15, k_z=3 and a single endogenous variable. Here, the effect size goes in the abscissa, whereas the power is shown in the ordinate. The upper bound corresponds to the power function for RB=0.15, whereas the lower bound is for RB=0.10. The results show that, as the effect size increases, so does the rejection rate. Therefore, our experiments show a loss of power as RB increases. The grey area between both bounds represents the loss of power. This result was expected, and in line with theory (Stock and Yogo, 2005). The rejection curve becomes steeper as RB increases, namely, the probability of rejecting the null hypothesis (Type I error) increases.

[16] A drawback of the two-stage version of the CF method is that the standard errors cannot be obtained directly from the information matrix, requiring alternative methods, such as the bootstrap.

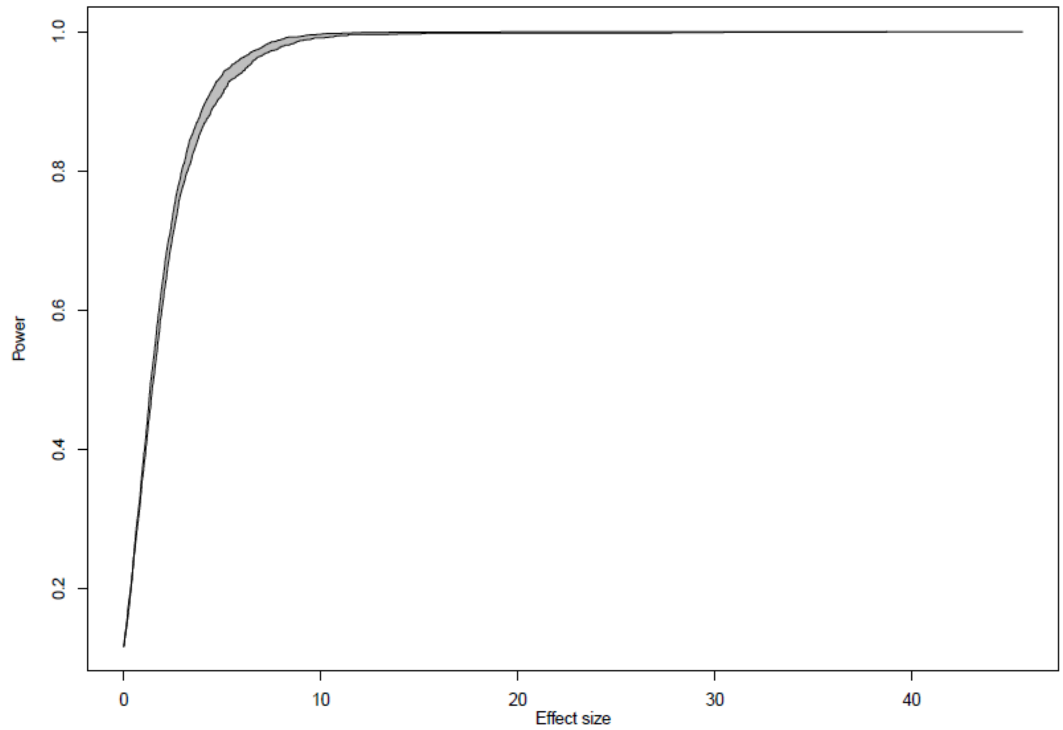

Figure 4-10. *Power function for RB criterion*

4.5. Conclusions and Future Research Directions

The effects of weak instruments have been extensively studied in linear models but not in-depth for DCM. Here, we address the problem contributing to bridge some research gaps and leave some open questions to be solved in this area. We believe that our findings will be useful in econometric modelling, especially in transport modelling, where this anomaly needs to be studied further. This research shows that the adverse effects of weak instruments in modelling cannot be neglected. As in linear models, also in DCM, weak instruments affect the estimation of consistent parameters.

This chapter provides three main contributions. First, we determined the CV from the F-statistic coming from the CF approach first stage for linear and DCM with a single endogenous regressor using the RB criterion. For this, we extended the results of Guevara and Navarro (2015), based on the findings of Stock and Yogo (2005) and improved by Skeels and Windmeijer (2018) for the case of the linear models. Our results are supported by the validation of our simulation approach for linear models. The critical values proposed for DCM are very similar to those found for linear models. Formally, we cannot reject the hypothesis that they are equal (Table 4-2 and 4-4). Our research's additional contribution is determining the critical value for k_z=1, for a single endogenous variable in the linear model and DCM (see Table 4-3 and 4-5). Our findings are in line with those reported by Skeels and Windmeijer (2018) in their Table A1 of Appendix D. These findings are significant and useful, overall, because in practice we often encounter difficulties in finding a sufficient number of instruments that fulfil the relevance condition.

Second, we show empirical evidence of the adverse effects of weak instruments on the DCM estimator's sample distribution. We found that, beyond reducing the degree of the correction of endogeneity, the use of weak instruments resulted in an increase of the variance of the estimators and a deformation of its sample distribution. This latter effect was not found for linear models, where the estimator's sample distribution retained the "bell" shape of a Normal distribution and did not deform. We also show that the critical values depend strongly on the relative bias that the modeller is willing to tolerate and the number of instruments (k_z), and that there is a loss of power as more relative bias is accepted.

The use of Monte Carlo simulation for these purposes represents our third contribution. We determined that the approach works properly by validating it for the case of linear models. The validation stage showed that the critical values for linear models are only not significantly different from those determined analytically from asymptotic distributions (Stock and Yogo, 2005; Skeels and Windmeijer, 2018). Thus, we consider that both approaches are appropriate.

Finally, we can propose several extensions for our work. The first is to use, instead of the relative bias, the size distortion of a Wald statistic as the measure to define the presence of weak instruments, following the work of Stock and Yogo (2005) for linear models. This type of analysis would allow controlling for the impact of weak instruments in inference but must surpass the challenge suggested by our empirical results, that the distributions seem to be no longer (and can be far from) Normal. Another extension would be to study the case of multiple endogenous regressors, where we can speculate that the statistic of Cragg and Donald (1993) may be more appropriate than the F-statistic, as recommended by Stock and Yogo (2005) for linear models. A third and final line of investigation in this topic, is related to the systematic analysis of the impact of considering different types of discrete choice models and functional forms in the analysis of the weak instrument problem.

5. Characterizing the Impact of Discrete Indicators to Correct for Endogeneity in Discrete Choice Models

5.1. Introduction

In this chapter we consider the MIS method for correcting endogeneity in DCM. The MIS method is based on the use of indicators, which often come from surveys designed for knowing respondents' attitudes and/or perceptions about their decision making (Bahamonde-Birke *et al*., 2017).

To do our tests we will use a purposely designed stated preference (SP) survey and Monte Carlo simulation. The SP survey was designed for the context of departure time choice, considering the main explanatory variables in this modelling context; that is, travel time, cost, variability of travel time and schedule delay (Arellana *et al*., 2012), where the last variable followed the scheduling model of Small (1982). The Monte Carlo experiments were designed to test the effect of: (*i*) different criteria for the discretization of latent continuous indicators, (*ii*) the sample size and (*iii*) the distribution of the indicator.

The rest of the chapter is organised as follows. Section 5.2 details the use of the MIS method in the case of DCM. The Monte Carlo experiment to assess the performance of discrete indicators using the MIS method is described in Section 5.3. In Section 5.4, we show the application of the MIS method for a departure time choice model estimated from SP data. In the final section we discuss the main findings, conclusions, and future research directions.

5.2. The MIS Approach in Discrete Choice Modelling

The MIS approach was initially introduced by Wooldridge (2010) for linear models, and later Guevara and Polanco (2016) extended it to DCM. As discussed above, the MIS uses indicators to correct for endogeneity. For explanatory purposes, let us consider the DCM represented by the utility function (U_{in}) in (5.1):

$$U_{in} = ASC_i + \beta_t t_{in} + \beta_c c_{in} + \beta_{sd} sd_{in} + e_{in}, \qquad (5.1)$$

where ASC_i is an alternative specific constant for alternative *i*, β_t, β_c, and β_{sd} are parameters to be estimated, t_{in} (time), c_{in} (cost), and sd_{in} (schedule delay) are the model's explanatory variables, and e_{in} is an exogenous error term; the subscript *n* represents the individual. For explanatory purposes, we will assume that t_{in} and c_{in} represent a set of known (measurable) attributes whereas the variable sd_{in} is both unknown to the modeller and correlated with t_{in}. Therefore, the modeller's specification would be as follows:

$$U_{in} = ASC_i + \beta_t t_{in} + \beta_c c_{in} + \varepsilon_{in}, \qquad (5.2)$$

where the error term ε_{in} contains both $\beta_{sd} sd_{in}$ and e_{in}. Thus, as the error term ε_{in} is correlated with t_{in} in (5.2) through sd_{in}, the DCM will suffer from endogeneity as a result of omitting sd_{in}; therefore, by definition, t_{in} is endogenous. Now, let us suppose that the variable sd_{in} and error terms ϵ_{in} can explain two indicators (I_{1in} and I_{2in}), as shown in (5.3) and (5.4):

$$I_{1in} = \alpha_1 + \alpha_{1sd} sd_{in} + \epsilon_{1in} \tag{5.3}$$

$$I_{2in} = \alpha_2 + \alpha_{2sd} sd_{in} + \epsilon_{2in} \tag{5.4}$$

We also assume that in this case the pairs of variables (sd_{in}, ϵ_{1in}), (t_{in}, ϵ_{1in}), (sd_{in}, ϵ_{2in}), (t_{in}, ϵ_{2in}), and (ϵ_{1in}, ϵ_{2in}) are mutually independent and that the coefficients α_{1sd} and α_{2sd} are not null; α_1 and α_2 are intercepts to be estimated.

Although sd_{in} in (5.2) is unknown to the researcher, for explanatory purposes we need to represent its effect on the model, and also to correct for the endogeneity resulting from its omission. For this, we can rewrite sd_{in} as a function of I_{1in} using (5.3) and by defining $\theta_{sd} = \frac{\beta_{sd}}{\alpha_{1sd}}$, the new expression for the utility function is given by (5.5):

$$U_{in} = ASC_i + \beta_t t_{in} + \beta_c c_{in} + \theta_{sd} I_{1in} \underbrace{-\theta_{sd}(\alpha_1 + \epsilon_{1in}) + e_{in}}_{\omega_{in}} \tag{5.5}$$

The model in (5.5) tries to correct for the endogeneity caused by omitting sd_{in} in (5.2), by including an indicator of it; however, the modified model still suffers from a different source of endogeneity because the term I_{1in} is correlated with ω_{in} through ϵ_{1in}.

The variable t_{in} is not endogenous in (5.5) because sd_{in} is no longer part of the error term; therefore, the only endogenous variable in (5.5) is I_{1in}, and to correct for endogeneity we need another instrumental variable. This may come from the second indicator I_{2in}, which by construction is correlated with I_{1in} only through sd_{in}, but it is independent of the error term ω_{in} in (5.5).

In linear models, this final correction for the MIS method is performed through TSLS (Wooldridge, 2010), but in DCM this is done using the CF method (Guevara and Polanco, 2016).

Thus, to apply the MIS method at least two indicators are needed for each endogenous variable considered. The MIS method uses one indicator to account for the omitted variable and the other as an instrument of the first one in the corrected model. Both effects are represented in (5.3) and (5.4). Where it can be noted that the variable omitted (sd_{in}) correlates to both indicators, and then in this way, I_{2in} can be used as an instrument of I_{1in} and vice versa.

In practice, for the endogenous DCM model in this example (5.2), the MIS method could be applied in two-stages as follows:

(i) apply an ordinary least squares (OLS) regression to I_{1in} on I_{2in}, t_{in} and c_{in}, and obtain the residuals $\hat{\delta}_{in}$. Both t_{in} and c_{in} must be included in this auxiliary regression because they are exogenous in the model (5.5)

(ii) estimate the DCM considering $\hat{\delta}_{in}$ and I_{1in} within the utility function.

Therefore, the DCM corrected for endogeneity using the MIS method (Guevara and Polanco, 2016) in two-stages would be as shown in (5.6) and (5.7):

$$I_{1in} = \alpha_t t_{in} + \alpha_c c_{in} + \alpha_{I_{2in}} I_{2in} + \delta_{in} \xrightarrow{OLS} \hat{\delta}_{in} = I_{1in} - \hat{I}_{1in} \quad (5.6)$$

$$U_{in} = \widehat{ASC}_i + \hat{\beta}_t t_{in} + \hat{\beta}_c c_{in} + \hat{\beta}_{I_{1in}} I_{1in} + \hat{\beta}_{\hat{\delta}} \hat{\delta}_{in} + \tilde{\varepsilon}_{in} \quad (5.7)$$

It can be noted that in (5.7) the term $\hat{\beta}_{\hat{\delta}} \hat{\delta}_{in} + \tilde{\varepsilon}_{in}$ is an orthogonal decomposition of ω_{in} in (5.5), in which the term $\hat{\beta}_{\hat{\delta}} \hat{\delta}_{in}$ captures the endogenous effect of ω_{in} that was present in (5.5). In this way, no term in (5.7) is correlated with the error term $\tilde{\varepsilon}_{in}$, and therefore, the endogeneity problem is accounted for.

5.3. A Departure Time SP Experiment to Assess the Performance of Discrete Indicators Using the MIS Method

In this section, we describe the MIS method's application to a databank coming from a specially designed SP survey in the context of departure time choice. The idea is to illustrate how, in practical terms, the use of continuous versus discrete indicators affects the results when using real data. We only provide a general description of the databank, limited to the main issues relevant to applying the MIS method.

Our SP experiment considered a survey where respondents were asked to report, first, their socioeconomic characteristics and information associated with their weekly trips by car to work. Regarding the latter, respondents had to report the usual travel time (*t*), any additional travel time (*v*) experienced in any of the trips during the week, and the usual departure time. This itinerary was considered as the *current* alternative. This information was also used as pivot to build two additional itineraries (one with an early departure time and another with a late departure time), that were eventually presented to respondents as alternatives in four hypothetical scenarios. To introduce a variable associated with the cost of travelling, we asked respondents to consider the existence of an urban toll, which depended on their trip departure times.

Departure time choice is typically modelled using the scheduling model formulated by Small (1982). Our questionnaire considered also asking for the preferred arrival time (PAT) to the respondents' jobs. As we knew the arrival time to the job (AT), because it is the departure time plus the travel time reported by the respondent, we were able to estimate the schedule delay early (sde) and the schedule delay late (sdl) attributes, as follows:

$$sde = Max\{(PAT) - (AT), 0\} \quad (5.8)$$

$$sdl = Max\{(AT) - (PAT), 0\} \quad (5.9)$$

Note that the values of (5.8) or (5.9) are either positive or zero. This implies that the individual will not experience disutility from rescheduling (Arellana *et al.*, 2012; Thorhauge *et al.*, 2016).

As our objective was to use the MIS method to correct for endogeneity, two indicators were collected for each set of itineraries shown. This part of the survey considered showing some (not all, to reduce the cognitive load) of the itineraries later presented in the choice tasks and asking the respondents to grade them using the two following indicators:

1) How pleasant is this itinerary for you?
2) How convenient is this departure time for you?

The graded itineraries did not include the cost variable, because the indicators were designed to capture the effect of schedule differences only. To complete the information for those itineraries that were not graded (to reduce the cognitive load), we used the multiple imputation approach proposed by Gopalakrishnan *et al.* (2020).

The hypothetical scenarios shown to respondents in the SP experiment, were generated using a D-efficient design (Rose and Bliemer, 2008), considering the principles of level balance and minimal overlap (Zwerina *et al*., 2005). We contacted a convenience sample, via the internet, and used the Qualtrics software to implement the survey. Each respondent was confronted with four hypothetical scenarios with three itineraries, where they had to choose the preferred one. With the data gathered, we estimated several models, some of which were corrected for endogeneity due to omitted attributes with the MIS method.

To study the impact of the discreteness of the indicators in the quality of the correction, we used three different scales for them, which were assigned at random to respondents. Thanks to the randomness, any difference in the quality of the correction attained in each case may be attributable to the degree of discreetness considered. The scale used were (*i*) from 0 to 5 (without decimals), with probability 50%; (*ii*) from 0 to 5 (including one decimal), with probability 25%; and (*iii*) from 0 to 100 (without decimals), with probability 25%. With scale (*i*) we could guarantee a purely discrete grade, whereas with scales (*ii*) and (*iii*) we could secure a more continuous grade. We did some preliminary runs to determine the performance of each grading scales for correcting endogeneity using the MIS method and found some numerical problems when using grading scale (*iii*). For this reason, we later scaled the data obtained from (*iii*), to convert it to the scale from 0 to 5 (including one decimal).

After data cleaning, the total number of valid choices recorded was 437 pseudo-observations graded with discrete indicators and 476 pseudo-observations graded with continuous indicators. The survey was applied in the five main cities of Colombia (Bogotá, Cali, Medellín, Bucaramanga and Barranquilla).

Our methodological contribution is related with the performance of the two types of indicators, discrete and continuous, to correct for endogeneity using the MIS method in this context. To the best of our knowledge, no research has reported indicators to correct for endogeneity in a time-of-day choice modelling context.

The estimated models were of Multinomial Logit (MNL) type with linear utilities. As future research we propose to correct for endogeneity using a more complex specification such as a Mixed Logit model with an error component to treat the pseudo panel nature of the SP data. The estimated models were the following:

1) Benchmark model: Containing all the explanatory variables considered in the SP experiment, with representative utility given by (5.10):

$$V_{in} = ASC_i + \beta_t t_{in} + \beta_c c_{in} + \beta_{sd} sd_{in} + \beta_v v_{in} \qquad (5.10)$$

Note that the parameter β_{sd} is specified as either β_{sde} or β_{sdl} for the early or late itineraries, respectively. The current itinerary is taken as reference (*ASC* fixed to zero) and does not have schedule delay. This model does not suffer from endogeneity since it considers all the attributes faced by the respondent and is therefore used as benchmark.

2) Endogenous model: As above but excluding sd_{in} and v_{in} in (5.10).

3) MIS-1 model: This is a model corrected for endogeneity and estimated including the variable $\hat{\delta}_{in}$ (coming from the first stage as explained above) and an indicator (I_{in}^1) within the utility function that can be discrete or continuous. The first and second stages of the MIS method follow expressions (5.11) and (5.12), respectively. Note that, as explained above, indicator 2 is used as an instrument for indicator 1.

$$I_{in}^1 = \alpha_1 + \alpha_t t_{in} + \alpha_c c_{in} + \alpha_{CI_{in}^2} I_{in}^2 + \delta_{in} \quad \xrightarrow{OLS} \quad \hat{\delta}_{in} = I_{in}^1 - \hat{I}_{in}^1 \qquad (5.11)$$

$$V_{in} = \widehat{ASC}_i + \hat{\beta}_t t_{in} + \hat{\beta}_c c_{in} + \hat{\beta}_{CI_{in}^1} I_{in}^1 + \hat{\beta}_{\hat{\delta}} \hat{\delta}_{in} \qquad (5.12)$$

4) MIS-2 model: Analogous to MIS-1, but in this case indicator 1 was used as an instrument of indicator 2.

$$I_{in}^2 = \alpha_1 + \alpha_t t_{in} + \alpha_c c_{in} + \alpha_{CI_{in}^1} CI_{in}^1 + \delta_{in} \quad \xrightarrow{OLS} \quad \hat{\delta}_{in} = I_{in}^2 - \hat{I}_{in}^2 \qquad (5.13)$$

$$V_{in} = \widehat{ASC}_i + \hat{\beta}_t t_{in} + \hat{\beta}_c c_{in} + \hat{\beta}_{CI_{in}^2} I_{in}^2 + \hat{\beta}_{\hat{\delta}} \hat{\delta}_{in} \qquad (5.14)$$

Table 5-1 presents the loglikelihood $l(\theta)$ and subjective values of time, SVT = $\left(\frac{\widehat{\beta}_t}{\widehat{\beta}_c}\right)$, for the various models estimated. We show the results for SVT[17] because the correction of endogeneity implies a change of scale (Guevara and Ben-Akiva, 2012); therefore, we can only analyse the ratios of the estimated coefficients to assess the phenomena under study. Besides, SVT is an important measure in the planning and social evaluation of transport projects (Ortúzar and Willumsen, 2011).

Databank	**Sample Size**	**Model**	$l(\boldsymbol{\theta})$	**SVT (COP/min)**
Discrete	437	Endogenous	-341.18	472.27
		Benchmark	-331.74	136.55
		MIS-1	-313.16	86.66
		MIS-2	-313.16	224.64
Continuous	476	Endogenous	-362.82	351.39
		Benchmark	-346.98	140.26
		MIS-1	-333.27	116.81
		MIS-2	-333.27	138.41

Table 5-1. *Summary of the $l(\theta)$ and SVT $\left(\frac{\widehat{\beta}_t}{\widehat{\beta}_c}\right)$ from the estimated models*

Further, $l(\theta)$ and SVT for the MIS-1 and MIS-2 models correspond to the mean for 100 simulations. This is because – as mentioned above - to deal with the missing indicators we used a multiple imputation process formulated by Gopalakrishnan *et al* (2020), and this process involved simulating the possible imputation values. These experiments were repeated with 100 random samples to avoid sampling bias.

The SVT estimated from the Benchmark models are similar to the values reported in the literature (Deloitte Consulting SLU, 2017). We checked that the sign (negative) of the estimators for t_{in}, c_{in}, v_{in}, and sd_{in} for the Benchmark models were as expected according to microeconomic theory. They also fulfilled the condition $\hat{\beta}_{sdl} < \hat{\beta}_{sde} < 0$ as stated in the literature (Arellana *et al.*, 2012; Koster and Verhoef, 2012; Thorhauge *et al.*, 2016). This means that people care more about being late instead of early. Note that the SVT achieved from the Benchmark models with both databanks are similar (136.55 and 140.26 COP/min).

The results in Table 5-1 show that the Endogenous models have clearly the worst performance. They show bias up to 246% in the estimates of the SVT in comparison with the Benchmark models. Given that the Endogenous models are a restricted version of the Benchmark models (5.10), it is possible to apply the likelihood ratio (LR) test (Ortúzar and Willumsen 2011, page 281) to compare them.

LR is asymptotically distributed χ_r^2 with *r* degrees of freedom, where *r* is the number of linear restrictions required to transform the more general model into the restricted version. In our case, r = 2 (because the restrictions are that both β_{sde} and β_{sdl} are zero), therefore, the LR test for both databanks are $LR = -2(-341.18 + 331.74) = 18.88$ and $LR = -2(-362.82 + 346.98) = 31.68$. These values must be compared with the critical value

[17] SVT is estimated in Colombian Pesos (COP)/minute. 1 US dollar is 3500 COP, approximately.

for two degrees of freedom at the 95% confidence level ($\chi_2^2 = 5.99$). As LR > χ_2^2 (in both cases), the null hypothesis of model equality is confidently rejected, and we can conclude that the Benchmark models are indeed superior.

On the other hand, the models corrected using the MIS method cannot be compared with the Endogenous or Benchmark model in terms of the LR test, because they are not a restricted version of them. Notwithstanding, the results in Table 5-1 show that the SVT reached from models MIS-1 and MIS-2 (116.81 COP/min and 138.41 COP/min, respectively) for the case with continuous indicators are closer to that of the Benchmark model. In this case, the bias attained from these estimates is 16.7% and 1.3%, respectively.

On the contrary, the SVT estimated from the databank with discrete indicators for MIS-1 and MIS-2 (86.66 and 224.64 COP/min, respectively) imply bias of 36.5% and 64.5% (respectively) in comparison with the SVT from the Benchmark model (136.55 COP/min). Note, also, that the difference between the SVT estimated from the databank with continuous indicators is only 21.6 COP/min (138.41 - 116.81), while the difference in the case of the discrete indicators is unusually large: 138.2 COP/min). Therefore, and in line with our initial hypothesis about endogeneity correction with the MIS method, using continuous indicators indeed performs better than using discrete indicators.

We also checked the parameter estimates and the MIS-1 and MIS-2 models' performance for the 100 replications needed to apply the multiple imputation method. First, the sign of the estimators for t_{in} and c_{in} was always as expected (negative) according to microeconomic theory. Second, when the continuous version of I_{in}^1 (*How pleasant is this itinerary for you?*) was included as an explanatory variable within the utility function (MIS-1 model), it always obtained a positive sign. This means that respondents tended to prefer itineraries that they graded better. Besides, the coefficient of $\hat{\delta}_{in}$ was significantly different from zero for all imputations, implying that the model including the continuous version of I_{in}^1 effectively suffered from endogeneity (Rivers and Vuong 1988).

On the other hand, when the continuous version of I_{in}^2 (*How convenient is this departure time for you?*) was included as an explanatory variable within the utility function (MIS-2 model), it showed the same performance, in terms of signs, as I_{in}^1. As for this case the coefficient of $\hat{\delta}_{in}$ was also significantly different from zero for all imputations, the presence of endogeneity is also evident in model MIS-2 (Rivers and Vuong 1988). Finally, the results in Table 5-1 show that the continuous versions of I_{in}^1 and I_{in}^2 have the same performance in terms of $l(\theta)$; however, the SVT are closer to the Benchmark model when I_{in}^1 was used as an instrument of I_{in}^2. This may be due to the fact that the question stated in I_{in}^2 captures better the effect of the variability of travel time and schedule delay in our SP experiment than that in I_{in}^1.

5.4. A Monte Carlo Experiment to Assess the Performance of Discrete Indicators Using the MIS Method

Our Monte Carlo experiment focused on three aspects: distribution of the indicator, sample size, and discretization process. This last aspect was considered relevant because, as we explain below, we assumed that the discrete indicators were obtained from continuous indicators; therefore, the discretization process should be explicit.

Our simulation experiment tried to emulate the same choice tasks of the departure time SP survey described in Section 5.3. For this, we generated data and estimated a DCM, following the scheduling model formulated by Small (1982), with three departure time choice alternatives: the current itinerary, an early itinerary and a late itinerary. The experiment was repeated 100 times to build a sampling distribution of the estimators (which becomes complicated in two stages methods) and to properly consider circumstantial effects.

We considered a DCM represented by the following utility function for alternative *i* and individual *n*:

$$U_{in} = ASC_i + \beta_t t_{in} + \beta_c c_{in} + \beta_{sd} sd_{in} + \beta_v v_{in} + e_{in} \qquad (5.15)$$

where, c_{in}, t_{in}, v_{in} and sd_{in} emulate the same explanatory variables of the model used in the SP experiment explained in the previous section, e_{in} is an error term that distributes Extreme Value Type I, and ASC_i is the alternative specific constant of alternative i. The parameter β_{sd} in (5.15) can be specified as either β_{sde} or β_{sdl} for the early or late itineraries, respectively. Again, the current itinerary does not have a schedule delay. The explanatory variables c_{in}, t_{in}, v_{in} and sd_{in} were generated using three distributions: Normal, Uniform and Exponential.

For simulation purposes, and without loss of generality, we assumed that the values of the parameters in (5.15) were as follows: $\beta_t = -4$, $\beta_c = -2$, and $\beta_v = -1$. The *ASC* for the early and late itineraries had values of -0.5 and -1.0, respectively. Finally, following the findings of Thorhauge *et al.* (2016), we defined the schedule delay early parameter $(\beta_{sde} = -1)$ with a smaller absolute value than the schedule delay late parameter $(\beta_{sdl} = -3)$.

To replicate the correlation among the SP survey databank variables (c_{in}, t_{in}, v_{in} and sd_{in}), in the simulated experiment we used the copulas approach[18] and the covariance matrix from SP survey databank to simulate correlated random variables. On the other hand, two appropriate indicators were generated for each itinerary, using equations (5.16) and (5.17) below. As these indicators are computed from the variables sd_{in}, v_{in} and the error term ϵ_{in}, they have a continuous nature and will be labelled as *CI*, where the superscripts 1 and 2

[18] The copulas are multivariate functions that join or "*couple*" two or more univariate distribution functions to construct continuous multivariate distribution functions. The copula represents a convenient parametric way to model the dependency structure on joint distributions of random variables, particularly for a set of random variables (Nelsen, 1999).

identify the indicator. To ensure consistency in the distribution of *CI* in (5.16) and (5.17), we forced sd_{in}, v_{in} and ϵ_{in} to have the same distribution. For example, if sd_{in}, v_{in} and ϵ_{in} distribute Normal, then *CI* distributes Normal.

$$CI_i^1 = \alpha_i^1 + \alpha_{sd}^1 sd_{in} + \alpha_v^1 v_{in} + \epsilon_{in}^1 \quad (5.16)$$

$$CI_i^2 = \alpha_i^2 + \alpha_{sd}^2 sd_{in} + \alpha_v^2 v_{in} + \epsilon_{in}^2 \quad (5.17)$$

Here, α_i are intercepts of the linear equations; they take the value of 1 for the early and current itineraries, and -1 for the late itinerary. The selection of these values is consistent with the fact that the early itinerary tends to be graded better than the late itinerary. Finally, the remaining parameters in (5.16) and (5.17) had the same value: $\alpha_{sd} = \alpha_v = -2$. This was done because in the first stage regression of the MIS method with the real dataset, these parameters turned out to be similar.

On the basis of the continuous indicators above, discrete indicators (DI_i^1 and DI_i^2) for itinerary *i* were generated using two discretization criteria and four discretization algorithms; these were implemented trying to avoid changing the original data distribution (Gonzalez-Abril *et al.*, 2009). The idea was to keep the high interdependence between the discrete indicators (DI_i^1 and DI_i^2) and the attributes sd_{in}, v_{in} and ϵ_{in}. The six discretization methods are described as follow:

1) The discrete indicator is equal to the continuous indicator's value rounded to the nearest integer.

2) The continuous indicator is assigned to a quintile interval (0-20, 20-40, …) and the closest integer belonging to the mean percentile of each interval (10, 30, …) is used as the discrete indicator.

3) The discrete indicator is estimated using the *minimum description length principle* (MDLP) algorithm (Fayyad and Irani, 1993).

4) The discrete indicator is estimated using the *class-attribute interdependence maximization* (CAIM) algorithm (Kurgan and Cios, 2004).

5) The discrete indicator is estimated using the *class-attribute contingency coefficient* (CACC) algorithm (Tsai *et al*., 2008), and

6) The discrete indicator is estimated using the *Ameva* algorithm (Gonzalez-Abril *et al*., 2009).

Detailed descriptions of algorithms (3) to (6) are given in the Appendix. Discretization algorithms have played an essential role in data mining and computer science (Tsai *et al*., 2008), as these areas often involve using continuous attributes that need to be implemented using categorical or discrete versions. So, many discretization algorithms have been proposed such as MDLP, CAIM, CACC and Ameva. We used these because they have been proved appropriate in the computer science field. Simulations were carried out for various sample sizes and for three distributions of the indicators, which depended on the distribution of the variables and error terms in (5.16) and (5.17). The aim was to analyze the effect of both conditions when correcting for endogeneity using the MIS method. Regarding sample size, we simulated three different situations: 5000, 2000 and 500 individuals. Finally, the variation of the distribution of the indicators was built by giving the variables sd_{in}, v_{in}, and the error term ϵ_{in} Normal, Uniform and Exponential distributions. In this way we could guarantee the following cases:

1) If sd_{in}, v_{in} and ϵ_{in} distribute Normal, the indicator distributes Normal.

2) If sd_{in}, v_{in} and ϵ_{in} distribute Uniform, the indicator distributes Triangular (if two terms are added) or Irving-Hall (if three terms are added).

3) If sd_{in}, v_{in} and ϵ_{in} distribute Exponential, the indicator distributes Gamma.

The observed variables in all cases were generated by means of the copulas approach (Nelsen, 1999), using the respective marginal distributions and considering a degree of correlation among them equal to the one observed in the SP experiment.

Finally, using the above data, choices were simulated and the databank from the Monte Carlo experiment was built. With this data, we estimated the following four models:

1) A *true* model, that considered all the explanatory variables described in (5.15). This model was used as benchmark.

2) An *endogenous* model estimated excluding the variables sd_{in} and v_{in} in (5.15) to cause endogeneity.

3) A *MIS DI* model that corrects the *endogenous* model including the variable $\hat{\delta}_{in}$ (coming from the first stage) and the discrete indicator DI_{in} within the utility function. The utility function of this model is shown in (5.18). Again, the indicator DI_{in}^2 was used as an instrument of DI_{in}^1.

$$U_{in} = \widehat{ASC}_i + \hat{\beta}_t t_{in} + \hat{\beta}_c c_{in} + \hat{\beta}_{DI_{in}^1} DI_{in}^1 + \hat{\beta}_{\hat{\delta}} \hat{\delta}_{in} + \tilde{\varepsilon}_{in} \qquad (5.18)$$

4) A *MIS CI* model corrected for endogeneity, estimated including $\hat{\delta}_{in}$ (coming from the first stage) and the continuous indicator CI_{in} within the utility function. The utility function of this model is shown in (5.19). This time, CI_{in}^2 was used as an instrument of CI_{in}^1.

$$U_{in} = \widehat{ASC}_i + \hat{\beta}_t t_{in} + \hat{\beta}_c c_{in} + \hat{\beta}_{DI_{in}^1} CI_{in}^1 + \hat{\beta}_{\hat{\delta}} \hat{\delta}_{in} + \tilde{\varepsilon}_{in} \qquad (5.19)$$

The four models described above were estimated for the different sample sizes, distribution of the indicators and discretization processes. As correcting for endogeneity in DCM implies a change of scale in the estimators (Guevara and Ben-Akiva, 2012), again we checked parameter ratios (VST = $\frac{\hat{\beta}_t}{\hat{\beta}_c}$) in the comparisons.

Table 5-2 shows the Monte Carlo experiment results for the various sample sizes, when the indicators distribute Normal and we use the first discretization criterium. The ratios among parameters ($\frac{\hat{\beta}_t}{\hat{\beta}_c}$) correspond to the mean of 100 replications. The p-value is for the null hypothesis (H_0) that the ratio of the parameters is equal to its population value. When H_0 cannot be rejected, the method is said to have properly corrected for endogeneity, and when H_0 is rejected, the method is said to have failed. For testing purposes, we took as critical a p-value of 0.05.

Sample size	**Model**	$\frac{\hat{\beta}_t}{\hat{\beta}_c}$	**Actual** $\frac{\beta_t}{\beta_c}$	**% Bias** $\frac{\hat{\beta}_t}{\hat{\beta}_c}$	**p-value**
5000	True	2.003	2.0	0.1	0.36
	Endogenous	1.932	2.0	3.4	<0.01
	MIS CI	2.003	2.0	0.1	0.38
	MIS DI	1.985	2.0	0.8	0.10
2000	True	1.998	2.0	0.1	0.39
	Endogenous	1.933	2.0	3.4	<0.01
	MIS CI	1.986	2.0	0.7	0.23
	MIS DI	1.995	2.0	0.2	0.38
500	True	2.008	2.0	0.4	0.37
	Endogenous	1.934	2.0	3.3	<0.01
	MIS CI	2.027	2.0	1.4	0.29
	MIS DI	2.051	2.0	2.5	0.10

Table 5-2. *Monte Carlo statistics for Normal indicators and discretization criterion 1 (rounded to the nearest integer)*

As can be seen, the true model always recovers the actual parameter ratio since its p-value is always far above the 5% threshold. On the contrary, the bias for the endogenous model is always significant. Under this simulation setting, where the indicators distribute Normal and the discretization criterium is simply rounding to the nearest integer, correcting for endogeneity using the MIS method works with both discrete and continuous indicators.

Table 5-3 shows the p-values corresponding to the ratio $\frac{\hat{\beta}_t}{\hat{\beta}_c}$ for the MIS DI model, for the three sample sizes, when varying the discretization criteria and the distribution of the indicators. Although not shown in the table, for the sake of space, it was always possible to recover the ratio between parameters for the true model and for the MIS method using continuous indicators. On the other hand, the ratio among parameters for the endogenous model failed to meet the target in several cases.

Discretization criterion	Indicator distribution	Sample size		
		5000	2000	500
1. Rounded to nearest integer	Normal	0.10	0.38	0.10
	Triangular	0.37	0.14	0.24
	Gamma	0.36	0.25	0.21
2. Percentiles	Normal	0.14	0.40	0.16
	Triangular	0.19	0.05	0.29
	Gamma	0.14	0.38	0.23
3. MDLP	Normal	0.29	0.19	0.06
	Triangular	<0.01	<0.01	<0.01
	Gamma	0.10	0.40	0.26
4. CAIM	Normal	<0.01	<0.01	<0.01
	Triangular	<0.01	<0.01	0.03
	Gamma	<0.01	<0.01	0.14
5. CACC	Normal	<0.01	<0.01	0.24
	Triangular	<0.01	<0.01	0.03
	Gamma	<0.01	<0.01	0.08
6. Ameva	Normal	<0.01	<0.01	0.02
	Triangular	<0.01	<0.01	0.03
	Gamma	<0.01	<0.01	<0.01

Table 5-3. *P-values corresponding to* $\frac{\hat{\beta}_t}{\hat{\beta}_c}$ *for the MIS DI model*

The p-values in green highlight the cases where the ratio $\frac{\hat{\beta}_t}{\hat{\beta}_c}$ for the MIS DI model works. On the contrary, the values in red show the cases where the correction failed under the 5% threshold. Finally, the values in black correspond to cases where the root-mean-squared error (RMSE) reached by the simulation was too large, yielding meaningless estimates of the p-value. This seems to be an effect of sample size because it only occurred for the smallest sample size (500 simulated individuals).

An interesting finding corresponds to the failure to correct for endogeneity when the CACC, CAIM and Ameva algorithms are used. These algorithms are widely used in machine learning and data mining, among others (Tsai *et al*., 2008; Gonzalez-Abril *et al*., 2009). Although their use guarantees correlation between the generated discrete indicator and the class attribute (i.e., the variables sd_{in} and v_{in}), in addition to preserving the distribution of the discrete indicator, this does not guarantee endogeneity correction.

Notwithstanding, the results for the MDLP algorithm are different. When the indicators distribute Normal or Gamma, the correction with discrete indicators works; however, this cannot be guaranteed when the indicator distributes triangular (Irvin-Hall). Although the MDLP algorithm does formally not preserve the continuous indicator distribution, we believe it worked better because the estimation of the discrete indicators was made over a wide range of values. This implies that, in practice, the approach tends to emulate a continuous distribution of values. This suggests that using more granular discrete indicators may be important to achieve a successful correction, although more research on this is needed to reach a firm conclusion.

To illustrate the impact of the different discretizing algorithms in the empirical distribution of the indicator, Figure 5-1 shows a scatter plot of matrices, with bivariate scatter plots below the diagonal, histograms on the diagonal, and the Pearson correlation index above the diagonal. The first column corresponds to the discrete indicator, the second and third represent the variables sd_{in} and v_{in}, respectively, and the last column represents the error term ϵ_{in} in (5.16) and (5.17). The results of MDLP, CAIM, CACC and Ameva are shown from left to right and top to bottom. The values above the main diagonal correspond to the correlation between the variables that are part of the scatter plot of matrices.

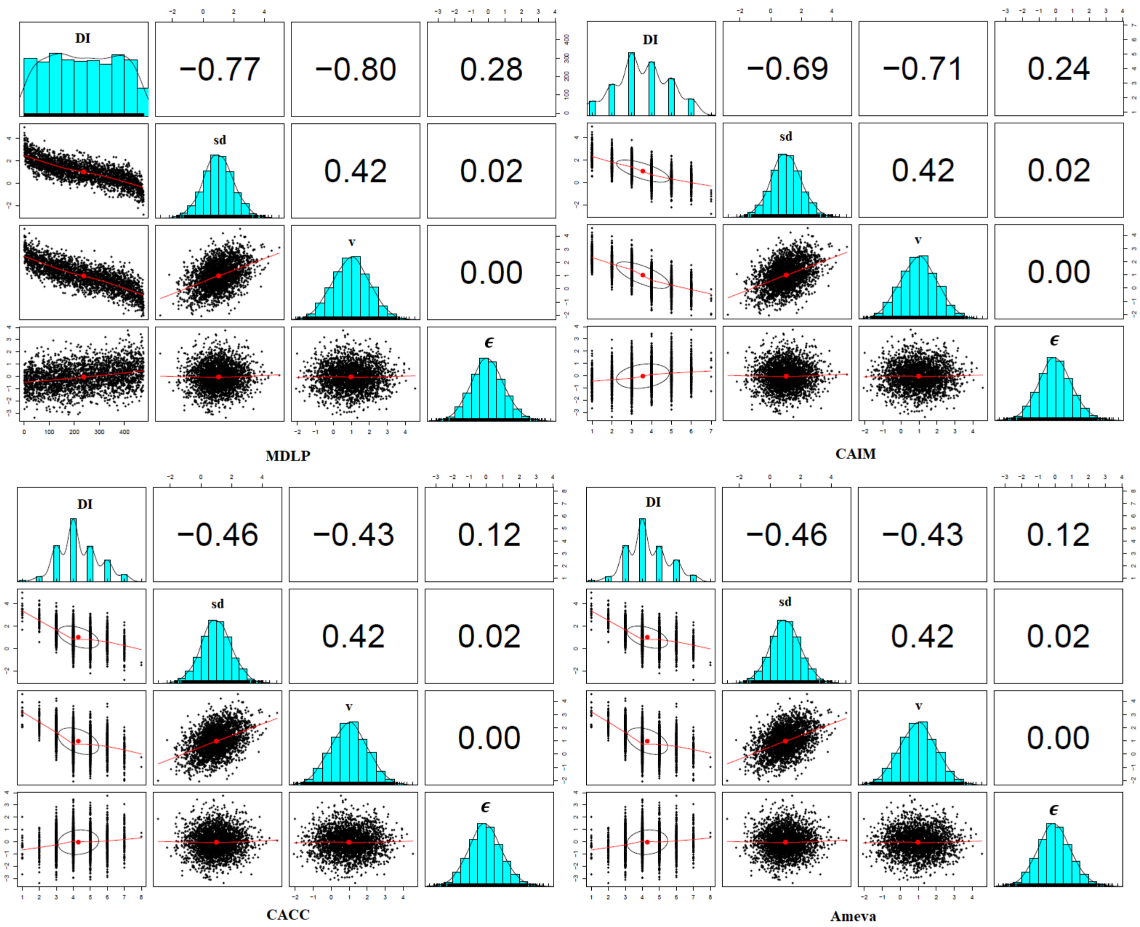

Figure 5-1. *Scatter plot for discrete indicator according to the algorithm*

Figure 5-1 shows that, as designed by the copula and discretization algorithm, the correlation between variables and indicators is different. For example, when the MDLP algorithm is used, the correlation between the indicator and sd_{in}, v_{in}, and ϵ_{in} achieved the highest values (-0.77, -0.80 and 0.28), in comparison with the other algorithms. Further, although all methods discretized the continuous indicator, the MDLP results appear substantially more granular (taking values between 0 and 500), while the other methods take values between 1 and 8. On the other hand, the MDLP results appear almost uniform, completely losing the continuous indicator's bell distributional shape, while the other methods do a better job at retaining it. So, given that MDLP was the only algorithm that worked in this experiment, we can speculate that the result is related to its granularity. Considering these findings, the correlation between the indicator and the class variables seems essential during the discretization process. In contrast, the change in the indicator distribution does not seem to be an aspect that affects the correction. Although these results are valuable, we consider that more research is needed to reach a firm conclusion in this sense.

5.5. Conclusions and Future Research Directions

Endogeneity is an anomaly that may yield inconsistent model parameters. We designed a Monte Carlo experiment and applied a SP survey, in the context of departure time choice, to examine the impact of using discrete indicators to correct for endogeneity with the MIS method in DCM. The reason was that, in practice, indicators tend to be discrete, but they should be continuous to be consistent with the mathematical derivation of the MIS method (Wooldridge, 2010).

From the Monte Carlo experiments, it was possible to show, first, that small sample sizes can lead to erroneous conclusions. As the mean square error increased, it erroneously led to conclude that the correction with discrete indicators worked. Second, the most straightforward criteria to produce discrete indicators, that is, *rounding the continuous value to the nearest integer* and *assign it to percentiles* (described in Section 5.4), worked properly. This may be due to the fact that these criteria preserve the correlation between the indicator and the class variables (sd_{in} and v_{in}). Similarly, we used four algorithms reported in the literature to discretize continuous indicators. We used these algorithms because they are typically used in the computer science field. Our findings show that the algorithm used to produce discrete indicators (from continuous ones) affects the endogeneity correction. Third, of the complex algorithms tested to perform the discretization process, only the MDLP algorithm was successful. This algorithm worked because the discrete indicators achieved from its application were substantially more granular than those of the other three methods, assimilating the distribution to a continuous shape. However, more research on this is needed to reach a strong conclusion.

On the other hand, using real data we were able to show that the correction with continuous indicators worked better than the correction with discrete indicators. In particular, we checked the parameter ratios (to bypass the scaling problem), finding that the subjective value of time (SVT) for the model corrected with continuous indicators was closer to that of the Benchmark model than those computed from the model corrected with discrete indicators. Notwithstanding, the SVT achieved with the endogenous model was much worse, having a bias of approximately 250% with respect to the benchmark model.

An interesting finding from this part of the research were the two indicators themselves, as they were able to capture the effect of the explanatory variables related to the schedule delay and travel time variability. This is a methodological contribution, as to the best of our knowledge, no previous research has suggested appropriate indicators to correct endogeneity under this modelling context.

Our findings also allow us to recommend that in future data collection, of this type, we should ask for indicators on a continuous scale, leaving aside the commonly used Likert scales. Besides, if discrete grades are used, respondents should be required to use wider ranges to achieve ratings throughout the entire scale. Our conclusions are valid for the conditions shown in our SP survey and the Monte Carlo experiment.

Finally, the findings and analyses performed in this chapter have several limitations that future research should address. For example, the impact of different types of discrete choice models and functional forms should be considered, as we only modelled the simplest case of a Logit model with fixed parameters. We also recommend exploring a mathematical derivation contemplating discrete distributions for the indicators.

6. Appendix

A detailed description of the four algorithms used in this research is reported below. The pseudocodes, description and notation used were extracted directly from the papers proposing the algorithms.

Class-attribute interdependence maximization (CAIM) algorithm

The CAIM algorithm was developed by Kurgan and Cios (2004). The CAIM criterion measures the dependency between a class variable *C* and a discretization variable *D* for attribute *F*, for a given quanta matrix[19] (see Table 6-1), and is defined as:

$$CAIM(C,D|F) = \frac{\sum_{r=1}^{n} \frac{max_r^2}{M_{+r}}}{n} \qquad (6.1)$$

Table 6-1. *2D quanta matrix for attribute F and discretization scheme D*

Class	Interval					Class Total
	$[d_0,d_1]$	…	$(d_{r-1},d_r]$	…	$(d_{n-1},d_n]$	
C_1	q_{11}	…	q_{1r}	…	q_{1n}	M_{1+}
:	:	…	:	…	:	:
C_i	q_{i1}	…	q_{ir}	…	q_{in}	M_{i+}
:	:	…	:	…	:	:
C_s	q_{S1}	…	q_{Sr}	…	q_{Sn}	M_{S+}
Interval Total	M_{+1}	…	M_{+r}	…	M_{+n}	

where *n* is the number of intervals, *r* iterates through all intervals (i.e., $r = 1, 2, \ldots, n$), max_r is the maximum value among all q_{ir} values (maximum value within the rth column of the quanta matrix), $i = 1, 2, \ldots, S$, M_{+r} is the total number of continuous values of attribute *F* that are within the interval $(d_{r-1}, d_r]$.

Kurgan and Cios (2004) describe the pseudocode of the CAIM algorithm. Given data consisting of M examples, S classes, and continuous attributes F_i. For every F_i do the following steps:

Step 1.

1.1 Find maximum (d_n) and minimum (d_0) values of F_i.

1.2 Form a set of all distinct values of F_i in ascending order, and initialize all possible interval boundaries B with minimum, maximum and all the midpoints of all the adjacent pairs in the set.

1.3 Set the initial discretization scheme as $D: \{[d_0, d_n]\}$ set GlobalCAIM=0.

[19] A quanta matrix is a two-dimensional frequency matrix where the class variable and the discretization variable of attribute F are treated as two random variables.

Step 2.

2.1 Initialize $k = 1$.

2.2 Tentatively add an inner boundary, which is not already in D, from B, and calculate the corresponding CAIM value.

2.3 After all the tentative additions have been tried, accept the one with the highest value of CAIM.

2.4 If (CAIM > GlobalCAIM or k < S) then update D with the accepted in Step 2.3 boundary and set GlobalCAIM=CAIM, else terminate.

2.5 Set $k = k + 1$ and go to 2.2.

Output: Discretization scheme D

Class-attribute contingency coefficient (CACC) algorithm

Tsai *et al*. (2008) developed this algorithm. They describe it as the pseudo-code shown in Figure 6-1. Given a dataset with *i* continuous attributes, *M* examples, and *S* target classes, for each attribute A_i, CACC first finds the maximum d_n and minimum d_0 of A_i in Line 4 and then forms a set of all distinct values of A_i in the ascending order in Line 5. As a result, all possible interval boundaries *B* with the minimum and the maximum, and all the midpoints of all the adjacent boundaries in the set are obtained in Lines 6 and 7.

```
1  Input:  Dataset with  i continuous attribute, M examples and S target classes;
2  Begin
3     For each continuous attribute A_i
4        Find the maximum d_n and the minimum d_0 values of  A_i;
5        Form a set of all distinct values of A in ascending order;
6        Initialize all possible interval boundaries B with the minimum and maximum
7        Calculate the midpoints of all the adjacent pairs in the set;
8        Set the initial discretization scheme as D: {[d_0,d_n]}and Globalcacc = 0;
9        Initialize k = 1;
10           For each  inner boundary  B which is not already in  scheme D,
11                Add it into D;
12                Calculate the corresponding cacc value;
13           Pick up the scheme D' with the highest cacc value;
14           If cacc  > Globalcacc or k < S  then
15                Replace D with D';
16                Globalcacc = cacc;
17                 k = k + 1;
18                Goto Line 10;
18           Else
19              D' = D;
20           End If
21    Output the Discretization scheme D'  with k intervals for continuous attribute A_i;
22 End
```

Figure 6-1. *The pseudo-code of CACC*

Then, CACC would iteratively partition the attribute A_i from Line 10 to Line 18. In the *kth* loop, CACC would compute for all possible cutting points to find the one with the maximum *cacc* value and then partition this attribute accordingly into *k* + 1 intervals. The *cacc* value is defined as follows:

$$cacc = \sqrt{\frac{y'}{y'+M'}} \quad (6.2)$$

where $y' = M\left[\left(\sum_{i=1}^{S}\sum_{r=1}^{n}\frac{q_{ir}^{2}}{M_{i+}M_{+r}}\right) - 1\right] / \log(n)$. To reduce the computational cost of the discretization, CACC also uses a greedy method (as in CAIM) to generate the sub-optimal discretization scheme. In other words, for every loop, CACC not only finds the best division point but also records a *Globalcacc* value.

If the generated *cacc* value in loop *k* + 1 is less than the *Globalcacc* obtained in loop *k*, CACC would terminate and output the discretization scheme. Besides, to generate a rational discrete result, such a greedy mechanism is ignored if the number of generated intervals is less than the number of target classes. Since the main framework of CACC is similar to that of CAIM, the complexity of CACC for discretizing a single attribute is still $O(m\log(m))$, where *m* is the number of distinct values of the discretized attribute.

Ameva algorithm

Gonzalez-Abril *et al*. (2009) developed this algorithm. They describe it as the pseudo-code shown in Figure 6-2. Let $X = \{X_1, \ldots, X_N\}$ be a training data set of a continuous attribute X of mixed-mode data such that each example x_i belongs to only one of the *l* classes of the class variable denoted by $\mathcal{I} = \{C_1, \ldots, C_l\}$. A continuous attribute discretization is a function $D: X \to \mathcal{I}$ which assigns a class $C_i \in \mathcal{I}$ to each value $x \in X$.

```
Input: Data consisting of N examples, ℓ classes, and
continuous variables Xi
For every Xi do:
Step 1: Initialization of the candidate interval
boundaries and the initial discretization scheme.
1.1 Find the maximum (dk) and minimum (do) values of Xi.
1.2 Form a set of all distinct values of Xi, in ascend-
ing order, and initialize all possible interval
boundaries B with the minimum, maximum and all the
midpoints of all the adjacent pairs in the set.
1.3 Set the initial discretization scheme to L:
{[d0, dk]}, set GlobalAmeva = 0.
Step2. Consecutive additions of a new boundary which
results in the locally highest value of the Ameva
criterion.
2.1 Initialize k = 1;
2.2 Tentatively add an inner boundary, which is not
already in L, from B, and calculate the corresponding
Ameva value.
2.3 After all the tentative additions have been tried,
accept the one with the highest value of Ameva.
2.4 If (Ameva > GlobalAmeva ) then update L with the
accepted boundary in step 2.3 and set
GlobalAmeva = Ameva , else terminate.
2.5 Set k = k + 1 and go to 2.2
Output: Discretization scheme L(k).
```

Figure 6-2. *The pseudo-code of Ameva*

Let us consider a discretization D which discretizes X into k discrete intervals $\{L_1, \dots, L_k\}$ where $L_1 = [(d_0, d_1]$ and $L_j = (d_{j-1}, d_j]$ for $j = 2, \dots, k$ such that d_k is the maximal value and d_0 is the minimal value of attribute X, and the values d_i are arranged in ascending order. Thus, a discretization variable is defined as $\mathcal{L}(k; X, \mathcal{I}) = \{L_1, \dots, L_k\}$ which verifies that, for all $x_i \in X$, a unique L_j exists such that $x_i \in L_j$ for $i = 1, \dots, N$ and $j = 1, \dots, k$. The discretization variable $\mathcal{L}(k; X, \mathcal{I})$ (denoted as $\mathcal{L}(k)$ for the sake of simplicity) of attribute X and the class variable $\mathcal{I}$ are treated from a descriptive point of view and Table 6-2 is drawn up where n_{ij} denotes the total number of continuous values belonging to the ith class that are within the jth interval, n_i is the total number of instances belonging to the ith class, and n_j is the total number of instances belonging to the jth interval for $i = 1, \dots, l$ and $j = 1, \dots, k$.

Table 6-2. *Contingency table for attribute X and discretization variable $\mathcal{L}(k)$*

$C_i \mid L_j$	L_1	...	L_j	...	L_k	n_j
C_1	n_{11}	...	n_{1j}	...	n_{1k}	n_1
...	...		...		...	...
C_i	n_{i1}	...	n_{ij}	...	n_{ik}	n_i
...	...		...		...	...
C_l	n_{l1}	...	n_{lj}	...	n_{lk}	n_l
n_j	n_1		n_j		n_k	N

Given discrete attributes $\mathcal{J}$ and $\mathcal{L}(k)$, the contingency coefficient, denoted by $\chi^2(k) \stackrel{\text{def}}{=} \chi^2(\mathcal{L}(k), \mathcal{I}|\mathrm{X})$, defined as:

$$\chi^2(k) = N\left(-1 + \sum_{i=1}^{l} \sum_{j=1}^{k} \frac{n_{ij}^2}{n_i n_j}\right) \qquad (6.3)$$

is considered. It is straightforward to prove that:

$$\max_{X, \mathcal{L}(k), \mathcal{I}} \chi^2(k) = N(min\{l, k\} - 1) \qquad (6.4)$$

Hence, the Ameva coefficient, $Ameva(k) \stackrel{\text{def}}{=} Ameva(\mathcal{L}(k), \mathcal{I}|\mathrm{X})$ is defined as follows:

$$Ameva(k) = \frac{\chi^2(k)}{k(l-1)} \qquad (6.5)$$

for $k, l \geq 2$. The Ameva criterion has the following properties:

- The minimum value of $Ameva(k)$ is 0 and when this value is achieved then both discrete attributes $\mathcal{J}$ and $\mathcal{L}(k)$ are statistically independent and vice versa.
- The maximum value of $Ameva(k)$ indicates the best correlation between the class labels and the discrete intervals. If $k \geq l$ then, for all $x \in C_i$ a unique j_0 exists such that x $x \in L_{j_0}$ (the remaining intervals $(k - l)$ have no elements); and if $k < l$ then, for all $x \in L_j$, a unique i_0 exists such that $x \in C_{i_0}$ (the remaining classes have no elements), that is, the highest value of the Ameva coefficient is achieved when all values within a particular interval belong to the same associated class for each interval.

- The aggregated value is divided by the number of intervals k, hence the criterion favours discretization schemes with the lowest number of intervals.
- To make a comparison with the CAIM coefficient, the aggregated value is divided by $l-1$.
- From (6.4), it follows that $Ameva_{max}(k) \stackrel{\text{def}}{=} max_{X,\mathcal{L}(k),\mathcal{J}}\ Ameva(k) = \frac{N(k-1)}{k(l-1)}$ if $k < l$ and $\frac{N}{k}$ otherwise. Hence, $Ameva_{max}(k)$ is an increasing function of k if $k \leq l$, and a decreasing function of k if $k > l$. Therefore, $max_{k\geq 2}\ Ameva_{max}(k) = Ameva_{max}(l)$, that is, the maximum of the Ameva coefficient is achieved in the optimal situation (all values of C_i are in a unique interval L_j and vice versa).

Minimum description length principle (MDLP) algorithm

Fayyad and Irani (1993) developed this algorithm. The MDLP criterion for the partition induced by a cut point *T* for a set *S* of *N* examples and for the continuous-valued attribute *A* is accepted if:

$$Gain(A,T;S) > \frac{\log_2(N-1)}{N} + \frac{\Delta(A,T;S)}{N} \qquad (6.6)$$

and it is rejected otherwise. Note that the quantities required to evaluate this criterion, namely the information entropy of S, S_1 and S_2 are computed by the cut point selection algorithm as part of cut point evaluation.

7. Conclusions

Endogeneity is an unavoidable anomaly in discrete choice models (DCM). This is serious, as DCM play a fundamental role in short and long-term transport planning and policy formulation (Ortúzar and Willumsen, 2011). In particular, mode choice (i.e., the third stage of the classic transport model) is typically modelled using DCM, and transport modelling applications are particularly susceptible to endogeneity problems: omitted attributes, measurement or specification errors, simultaneous determination and/or self-selection are common in this field. If the effects of endogeneity are not considered, the analysis can lead to wrong forecasts and conclusions (Guevara and Ben-Akiva, 2006), and to potentially faulty decision making.

We used the CF approach to address the first and second objectives of this Ph.D. thesis. With this, it was possible to correct the endogeneity of mode choice models at the strategic level using appropriate instrumental variables. Besides, its performance was tested in the forecasting stage and in designing a methodology based on Monte Carlo method simulations to detect weak instruments in linear models and DCM.

On the other hand, we used the MIS method as an alternative tool to address the thesis' third objective. Contrary to the CF approach, the MIS method uses indicators instead of instruments, which in theory should be easier to obtain as instruments must fulfil two requirements that are many times difficult to achieve. The CF approach and the MIS method can be considered two adequate methodologies to address each case's problem.

From the findings related with the first objective, developed in Chapters 2 and 3, this research provides the followings conclusions:

a) A framework that uses the CF method to correct for the endogeneity of mode choice models at the strategic level using appropriate instrumental variables. The instruments used were: (*i*) The average travel time of other origin-destination pairs that have a similar length than the origin and destination of the considered trip; (*ii*) the average travel cost of other origin-destination pairs that have a similar length than the origin and destination of the considered trip, and (*iii*) the network trip distance between the origin and the destination for each mode.
b) The confidence in strategic urban mode choice models based on level-of-service variables, such as travel cost and travel time, must be questioned. Our results show that the cost parameters could be more poorly estimated than the time parameters. This may be due to the fact that urban mode choice models at the strategic level may be affected by three sources of endogeneity: measurement errors, omitted variables and simultaneous estimation. We recommended: (*i*) to use instruments within the framework shown in this thesis to improve the estimations, and (*ii*) to focus the efforts in improving the way the cost variable is collected and measured in surveys, to achieve more consistent parameters during model estimation.

c) The effects of endogeneity in strategic urban mode choice models were quantified. Our findings show that the subjective value of time (SVT) was overestimated by 43% and 26% for private and public modes, respectively, in our case study. This fact may have a strong influence in the social evaluation of transport projects where the SVT is critical. We also looked at the impact on model elasticities, finding that these were underestimated. In particular, the Generalised Time elasticities showed underestimations of up to 33%, while the Cost/Income elasticities reached underestimations of up to 75%.

d) We provided a framework for correcting endogeneity, which was applied at the forecasting stage of supply-demand equilibration models when the endogenous variables changed over the years. A complex transport modelling process was emulated and affected by three different endogeneity sources that are common in strategic studies (measurement error, omitted variables and simultaneous estimation in a supply-demand equilibration context). In this setting, we compared three different approaches: (*i*) No endogeneity correction, which has been, so far, the only method used in practice, (*ii*) the *CF* approach proposed by Guevara and Ben-Akiva (2012) and (*iii*) a new proposal, the *CFU* approach. Forecasts were evaluated in terms of recovery of the *true* (simulated) travel times, and two goodness-of-fit indices, $E(l(\theta))$ and the $E(AIC)$, for future scenarios in 10 to 40 years ahead.

e) Monte Carlo simulation helped us to demonstrate that under fairly reasonable conditions, consistent with observed data, the new *CFU* proposal performed better than the other two approaches; in particular, it performed much better than the "*do nothing*" approach and marginally (but significantly) better than the more classical *CF* approach of Guevara and Ben-Akiva (2012). We showed that the adverse effects of endogeneity increase over the years, severely impacting the forecasts for future scenarios.

f) Our methodological findings suggest two important recommendations for practice. The first is to avoid forecasting with endogenous models as the problem is severe in the case studied. The second is that even when correcting for endogeneity, the residuals from the first stage of the *CF* approach for future scenarios should always be updated in forecasting. Thus, our new *CFU* approach is especially recommended to correct for endogeneity when discrete choice models are used in forecasting for strategic scenarios involving supply-demand equilibration.

From the findings reached in the case of the second objective developed in Chapter 4, we can provide the followings conclusions:

a) The effects of weak instruments have been extensively studied in linear models but not in-depth for DCM. Our findings are especially useful in transport modelling, but they can be used also in other areas such as road safety, marketing, spatial economics, tourism, urbanism, and environmental economics, where this anomaly needs to be studied further. We confirmed that the adverse effects of weak instruments in modelling cannot be neglected. As in linear models, weak instruments affect the estimation of consistent parameters in DCM.

b) Three main contributions can be summarized from our findings. First, we determined critical values from the F-statistic coming from the first stage of the *CF* approach for linear models and DCM, with a single endogenous regressor, using the relative bias criterion. Results are supported by the validation of our simulation approach in the case of linear models. The critical values proposed for DCM are very similar to those found for linear models; formally, we cannot reject the hypothesis that they are equal (Tables 4-2 and 4-4). Also, we were able to determine the critical value for k_z=1, for a single endogenous variable in linear models and DCM (see Tables 4-3 and 4-5). Our findings are in line with those reported by Skeels and Windmeijer (2018) in Table A1 of their Appendix D. These findings are significant and useful, overall, because in practice we often encounter difficulties in finding a sufficient number of instruments that fulfil the relevance condition.

   Second, we show empirical evidence about the adverse effects of weak instruments on the sample distribution of DCM estimators. We found that, beyond reducing the degree of endogeneity correction, the use of weak instruments results in an increase of the variance of the estimators and a deformation of their sample distribution. This latter effect was not found for linear models, where the estimator's sample distribution retained the "bell" shape of a Normal distribution and did not deform. We also showed that the critical values depend strongly on the relative bias that the modeller is willing to tolerate and the number of instruments (k_z), and that there is a loss of power as more relative bias is accepted.

   Third, the use of Monte Carlo simulation for these purposes represents our third contribution. We determined that the approach works properly by validating it for the case of linear models. The validation stage showed that the critical values for linear models are not significantly different from those determined analytically from asymptotic distributions (Stock and Yogo, 2005; Skeels and Windmeijer, 2018). Thus, we consider that both approaches are appropriate.

Finally, from the findings reached in the third objective developed in Chapter 5, we provide the followings conclusions:

a) We were able to characterize how, under certain conditions, the correction for endogeneity using the MIS method can be affected when discrete indicators are used instead of continuous indicators. From this point of view, a Monte Carlo experiment and an SP survey were built for the context of departure time choice. Both cases show the effects of each type of indicator on the endogeneity correction.
b) From the Monte Carlo experiments, it was possible to show that, first, small sample sizes can lead to erroneous conclusions because the mean square error increases, erroneously leading to conclude that correcting with discrete indicators works.

Second, the most straightforward criteria to produce discrete indicators, that is, *rounding the continuous values to their nearest integer* and *assigning them to percentiles* (described in Section 5.4) work properly. This can be due to the fact that these criteria preserve the correlation between the indicator and the class variables (sd_{in} and v_{in}). Third, we used four algorithms from the specialised literature in computer science to discretize continuous attributes finding, first, that they were not neutral in terms of endogeneity correction. In fact, only the MDLP algorithm was successful in this sense, and it worked because the discrete indicators it produced were substantially more granular, their distribution assimilating better to a continuous shape. However, more research on this is needed to reach a strong conclusion.

c) Using real data, we were able to show that correcting with continuous indicators had a better performance than with discrete indicators. We compared the SVT estimates from a benchmark model with SVT estimates achieved with the MIS correction. The SVT derived from the model with continuous indicators was much closer to that achieved with the model using discrete indicators. On the other hand, the SVT achieved with an endogenous model had a bias of approximately 250% with respect to the benchmark model. Therefore, we are able to suggest that if there is need to correct for endogeneity using the MIS method, indicators should be gathered on a continuous scale leaving aside the commonly used Likert scales. Finally, the two indicators considered in our application were able to capture the effect of the explanatory variables related with the schedule delay and travel time variability. This is a methodological contribution because both indicators can be recommended as appropriate for future research in this context. To the best of our knowledge, there is no previous research suggesting indicators to correct for endogeneity under this modelling context.

8. References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-723.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model. Econometrica **46**, 1193–1205.

Andrews, I., Stock, J.H. & Sun, L. (2019). Weak instruments in instrumental variables regression: theory and practice. *Annual Review of Economics* **11**, 727-753.

Angrist, J.D. & Pischke, J.S. (2009). *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton University Press, Princeton.

Anderson, J. & Hernandez, S. (2017). Roadway classifications and the accident injury severities of heavy-vehicle drivers. *Analytic Methods in Accident Research* **15**, 17-28.

Arellana, J., Daly, A., Hess, S., Ortúzar, J. de D. & Rizzi, L. I. (2012). Development of surveys for study of departure time choice: two-stage approach to efficient design. *Transportation Research Record* **2303**, 9-18.

Bahamonde-Birke, F.J., Kunert, U., Link, H. & Ortúzar, J. de D. (2017). About attitudes and perceptions: finding the proper way to consider latent variables in discrete choice models. *Transportation* **44**, 475-493.

Barnow, B., Glen, C., & Arthur, G. (1981). Selection on Observables. *Evaluation Studies Review Annual* **5**, 43-59.

Bass, F. (1969). A simultaneous equation regression study of advertising and sales of cigarettes. *Journal of Marketing Research* **6**, 291–300.

Bass, P., Donoso, P. & Munizaga, M. (2011). A model to assess public transport demand stability. *Transportation Research Part A: Policy and Practice* **45**, 755-764.

Ben-Akiva, M.E. & Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, Mass.

Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. & Polydoropoulou, A. (2002). Integration of choice and latent variable models. En H.S. Mahmassani (ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges.* Pergamon, Amsterdam.

Berry, S. & Jia, P. (2010). Tracing the woes: an empirical analysis of the airline industry. *American Economic Journal: Microeconomics* **2**, 1–43.

Berry, S., Levinsohn, J. & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica* **63**, 841-890.

Deloitte Consulting SLU (2017). Evaluación Socioeconómica para la Primera Línea de metro de Bogotá. Deloitte Consulting SLU, Bogotá.

Bolduc, D. & Alvarez-Daziano, R. (2009). On estimation of hybrid choice models. En S. Hess & A. Daly (eds.), *Choice Modelling: The State-of-the-Art and the State-of-Practice.* Emerald Group Publishing Limited, Bingley.

Bound, J., Jaeger, D.A. & Baker, R.M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* **90**, 443–450.

Bresnahan, T.F. (1997). The apple-cinnamon cheerios war: valuing new goods, identifying market power, and economic measurement. In T.F. Bresnahan & R.J. Gordon (eds.), *The Economics of New Goods.* NBER Studies in Income and Wealth Nº 58, The University of Chicago Press. Chicago.

Bun, M.J. & Harrison, T.D. (2018). OLS and IV estimation of regression models including endogenous interaction terms. *Econometric Reviews* **38**, 814-827.

Bureau of Public Roads (1964). *Traffic Assignment Manual*. Urban Planning Division, US Department of Commerce, Washington, DC.

Casey, B. (1989). Early retirement: the problems of “instrument substitution” and “cost shifting” and their implications for restructuring the process of retirement. En W. Schmähl (ed.), *Redefining the Process of Retirement*. Springer, Berlin.

Caussade, S., Ortúzar, J. de D., Rizzi, L.I. & Hensher, D.A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological* **39**, 621-640.

Chou, C.J. & Chen, P.C. (2014). Preferences and willingness to pay for green hotel attributes in tourist choice behaviour: the case of Taiwan. *Journal of Travel & Tourism Marketing* **31**, 937-957.

Citilabs (2016) Discover Cube 6.4 Tutorial. Citilabs, Inc., Sacramento.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Science*s. Lawrence Erlbaum Associates Publishers, New York.

Cragg, J.G. & Donald S.G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory* **9**, 222–40.

Cramer, J.S. (2007). Robustness of logit analysis: unobserved heterogeneity and mis-specified disturbances. *Oxford Bulletin of Economics and Statistics* **69**, 545-555.

Daly, A., Hess, S., Patruni, B., Potoglou, D. & Rohr, C. (2012). Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation* **39**, 267-297.

Davison, A.C. & Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge.

De Cea, J., Fernández, J.E., Dekock, V. & Soto, A. (2005) Solving network equilibrium problems on multimodal urban transportation networks with multiple user classes. *Transport Reviews* **35**, 293–317.

Dufour, J.M. & Wilde, J. (2018). Weak identification in Probit models with endogenous covariates. *AStA Advances in Statistical Analysis* **102**, 611-631.

Domencich, T. & McFadden, D. (1975). *Urban Travel Demand: A Behavioural Analysis*. North Holland, Amsterdam.

Fernández-Antolín, A., Guevara, C.A., De Lapparent, M. & Bierlaire, M. (2016). Correcting for endogeneity due to omitted attitudes: empirical assessment of a modified MIS method using RP mode choice data. *Journal of Choice Modelling* **20**, 1-15.

Gaudry, M.J., Jara-Diaz, S.R. & Ortúzar, J. de D. (1989). Value of time sensitivity to model specification. *Transportation Research Part B: Methodological* **23**, 151-158.

Gayle, P.G. (2004). Does price matter? Price and non-price competition in the airline industry. *North American Summer Meeting of the Econometric Society*. University of Minnesota, Minneapolis.

Glerum, A., Atasoy, B. & Bierlaire, M. (2014). Using semi-open questions to integrate perceptions in choice models. *Journal of Choice Modelling* **10**, 11-33.

Godfrey, L.G. (1999). Instrument relevance in multivariate linear models. *Review of Economics and Statistics* **81**, 550-552.

Gonzalez-Abril, L., Cuberos, F.J., Velasco, F. & Ortega, J.A. (2009). Ameva: an autonomous discretization algorithm. *Expert Systems with Applications* **36**, 5327-5332.

Gopalakrishnan, R., Guevara, C. A., & Ben-Akiva, M. (2020). Combining multiple imputation and control function methods to deal with missing data and endogeneity in discrete-choice models. *Transportation research part B: methodological* **142**, 45-57.

Granados, N., Gupta, A. & Kauffman, R.J. (2012). Online and off-line demand and price elasticities: evidence from the air travel industry. *Information Systems Research* **23**, 164-181.

Green, W.H. (1999). *Análisis Econométrico*. Prentice Hall, Madrid.

Guerrero, T.E., Guevara, C.A., Cherchi, E. & Ortúzar, J. de D. (2020). Addressing endogeneity in strategic urban mode choice models. *Transportation*, https://doi.org/10.1007/s11116-020-10122-y.

Guevara, C.A. (2010). *Endogeneity and Sampling of Alternatives in Spatial Choice Models.* Ph.D. Thesis, Department of Civil and Environmental Engineering, 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.

Guevara, C.A. & Ben-Akiva, M.E. (2006). Endogeneity in residential location choice models. *Transportation Research Record* **1977**, 60–66.

Guevara, C.A. & Ben-Akiva, M.E. (2010). Addressing endogeneity in discrete choice models: assessing control-function and latent-variable methods. En S. Hess & A. Daly (eds.), *Choice Modelling: The State-of-the-Art and The State-of-Practice*. Edward Elgar Publishers, Cheltenham.

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

Guevara, C.A. & Navarro, P. (2015). Detection of weak instruments when correcting for endogeneity in binary logit models. *Proceedings 14th International Conference on Travel Behaviour Research*, Windsor.

Guevara, C.A. & Polanco, D. (2016). Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. *Transportmetrica A: Transport Science* **12**, 458-478.

Guevara, C.A., Tirachini, A., Hurtubia, R. & Dekker, T. (2020). Correcting for endogeneity due to omitted crowding in public transport choice using the Multiple Indicator Solution (MIS) method. *Transportation Research Part A: Policy and Practice* **137**, 472-484.

Hardin, J.W. (2002). The robust variance estimator for two-stage models. *The Stata Journal* **2**, 253-266.

Hausman, J.A. (1978). Specification tests in econometrics. *Econometric*a **46**, 1251–1271.

Hausman, J.A. (1996). Valuation of new goods under perfect and imperfect competition. In T.F. Bresnahan & R.J. Gordon (eds.), *The Economics of New Goods*. NBER Studies in Income and Wealth Nº 58, The University of Chicago Press. Chicago.

Hausman, J.A., Leonard, G. & Zona, J.D. (1994). Competitive analysis with differentiated products. *Annals of Economics and Statistics* **34**, 159–180.

Heckman, J.J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica* **46**, 931–959.

Hess, S. & Beharry-Borg, N. (2012). Accounting for latent attitudes in willingness-to-pay studies: the case of coastal water quality improvements in Tobago. *Environmental and Resource Economics* **52**, 109-131.

Hsiao, C.Y. (2008). *Passenger Demand for Air Transportation in a Hub-and-Spoke Network*. Ph.D. Thesis, Department of Civil and Environmental Engineering, University of California at Berkeley.

Hurtubia, R. & Bierlaire, M. (2014). Estimation of bid functions for location choice and price modelling with a latent variable approach. *Networks and Spatial Economics* **14**, 47-65.

INRO (1996) EMME/2 *User's Manual*. INRO Inc., Montreal.

Instituto Nacional de Estadísticas (2013) Chile proyecciones y estimaciones de población. http://www.ine.cl/canales/chile_estadistico/demografia_y_vitales/demografia/demografia.php. Accessed July 24 2019.

Jensen, A.F., Cherchi, E. & Mabit, S.L. (2013). On the stability of preferences and attitudes before and after experiencing an electric vehicle. *Transportation Research Part D: Transport and Environment* **25**, 24-32.

Jensen, A.F., Cherchi, E., Mabit, S.L. and Ortúzar, J. de D. (2016). Predicting the potential market for electric vehicles. *Transportation Science* **51**, 427-440.

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

Kinal, T.W. (1980). The existence of moments of k-class estimators. *Econometrica* **49**, 241–249.

Koster, P., & Verhoef, E. T. (2012). A rank-dependent scheduling model. *Journal of Transport Economics and Policy* **46**, 123-138.

Lam, S.K., Ahearne, M., Hu, Y. & Schillewaert, N. (2010). Resistance to brand switching when a radically new brand is introduced: a social identity theory perspective. *Journal of Marketing* **74**, 128-146.

Lee, L. (1982). Specification error in multinomial logit models. *Journal of Econometrics* **20**, 197–209.

Lee, L. (1992). Amemiya's generalized least squares and tests of overidentification in simultaneous equation models with qualitative or limited dependent variables. *Econometric Reviews* **11**, 319–328.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology* **140**, 1-55.

Liu, H.X., He, X. & He, B. (2009). Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. *Networks and Spatial Economics* **9**, 485-503.

Louviere, J.J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T.A. & Waldman, D. (2005). Recent progress on endogeneity in choice modelling. *Marketing Letters* **16**, 255-265.

Lurkin, V., Garrow, L.A., Higgins, M.J., Newman, J.P. & Schyns, M. (2017). Accounting for price endogeneity in airline itinerary choice models: an application to Continental U.S. markets. *Transportation Research Part A: Policy and Practice* **100**, 228–246.

Magnusson, L.M. (2007). Weak instruments robust tests for limited dependent variable models. *Working Paper*, Department of Economics, Brown University.

Mariel, P., Hoyos, D., Artabe, A. & Guevara, C.A. (2018). A multiple indicator solution approach to endogeneity in discrete-choice models for environmental valuation. *Science of the Total Environment* **633**, 967-980.

Márquez, L., García, D.E. & Guarín, L.C. (2014). Funciones volumen-demora BPR y cónica en vías multicarriles de Bogotá. *Revista de Ingeniería* **41**, 30-39.

MathWorks. (2016). *MATLAB and Statistics Toolbox Release 2016b*. The MathWorks Inc., Natick.

Mumbower, S., Garrow, L.A. & Higgins, M.J. (2014). Estimating flight-level price elasticities using online airline data: a first step toward integrating pricing, demand, and revenue optimization. *Transportation Research Part A: Policy and Practice* **66**, 196–212.

Nelsen, R.B. (1999). *An Introduction to Copulas.* Springer, New York.

Nelson, C.R. & Startz, R. (1990a). Some further results on the exact small sample properties of the instrumental variables estimator. *Econometric*a **58**, 967–976.

Nelson, C.R. & Startz, R. (1990b). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business* **63**, 5125–5140.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**, 307–342.

Newey, W. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* **36**, 231–250.

Ortúzar, J. de D. & Willumsen, L.G. (2011). *Modelling Transport*. 4th ed. John Wiley & Sons, Chichester.

Park, S. & Gupta, S. (2009). Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. *Journal of Marketing Research* **46**, 531-542.

Petrin, A. & Train, K. (2002). Omitted product attributes in discrete choice models. Working Paper, Department of Economics, University of California, Berkeley.

Petrin, A. & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* **47**, 3-13.

Phillips, P.C. (1989). Partially identified econometric models. *Econometric Theory* **5**, 181-240.

R Development Core Team (2008). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Viena http://www.Rproject.org.

Raveau, S., Alvarez-Daziano, R., Yañez, M.F., Bolduc, D. & Ortúzar, J. de D. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record* **2156**, 131–139.

Rivers, D. & Vuong, Q. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* **39**, 347–366.

Rizzi, L.I. & Ortúzar, J. de D. (2006). Road safety valuation under a stated choice framework. *Journal of Transport Economics and Policy* **40**, 69-94.

Rose, J.M. & Bliemer, M.C.J. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews* **29**, 587-617.

Ross, A. & Willson, V.L. (2018). *Basic and Advanced Statistical Tests: Writing Results Sections and Creating Tables and Figures*. Sense Publishers, Rotterdam.

Rothenberg, T.J. (1984). Approximating the distributions of econometric estimators and test statistics. *Handbook of Econometrics* **2**, 881–935. North Holland, Amsterdam.

Ruud, P. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete models. *Econometrica* **51**, 225–228.

Sanderson, E. & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics* **190**, 212-221.

Sargan, J. (1958). The estimation of economic relationships using instrumental variables. *Econometrica* **26**, 393–415.

SECTRA (2013). Mediciones de Aforos de Tráfico y Perfiles de Carga en Servicios Troncales en el Gran Santiago. SECTRA, Santiago.

SECTRA (2014a). Actualización Diagnóstico del S.T.U. del Gran Valparaíso, Etapa I-SECTRA, Valparaiso.

SECTRA (2014b). Actualización y Recolección de Información del Sistema de Transporte Urbano, IX Etapa: Encuesta Origen Destino Santiago 2012, SECTRA, Santiago.

Shea, J. (1997). Instrument relevance in multivariate linear models: a simple measure. *Review of Economics and Statistics* **79**, 348–352.

Shin, H., Kim, D.K., Kho, S.Y. & Cho, S.H. (2020). Valuation of metro crowding considering heterogeneity of route choice behaviours. *Transportation Research Record*, https://doi.org/10.1177/0361198120948862

Skeels, C.L. & Windmeijer, F. (2018). On the Stock–Yogo tables. *Econometrics* **6**, 1-23.

Staiger, D. & Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica* **65**, 557 – 586.

Stern, S. (1996). Market definition and the returns to innovation: substitution patterns in pharmaceutical markets. Working Paper, Sloan School of Management, MIT.

Stock, J.H. & Yogo, M. (2005). Testing for weak instruments in linear IV regression. En D.W. Andrews & J.H. Stock (eds.), *Identification and Inference for Econometric Models, Essays in Honour of Thomas Rothenberg*. Cambridge University Press, Cambridge.

Terza, J.V. (2016). Simpler standard errors for two-stage optimization estimators. *The Stata Journal* **16**, 368-385.

Thorhauge, M., Cherchi, E., & Rich, J. (2016). How flexible is flexible? Accounting for the effect of rescheduling possibilities in choice of departure time for work trips. *Transportation Research Part A: Policy and Practice* **86**, 177-193.

Tirachini, A., Hensher, D.A. & Rose, J.M. (2013). Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice* **53**, 36–52.

Torres, I., Greene, M. & Ortúzar, J. de D. (2013). Valuation of housing and neighbourhood attributes for city centre location: a case study in Santiago. *Habitat International* **39**, 62-74.

Train, K.E. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press, Cambridge.

Train, K. & Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. En R. Scarpa & A. Alberini (eds.), *Applications of Simulation Methods in Environmental and Resource Economics*. Springer, Dordrecht.

Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass.

Varela, J.M., Börjesson, M. & Daly, A. (2018). Quantifying errors in travel time and cost by latent variables. *Transportation Research Part B: Methodological* **117**, 520-541.

Varotto, S.F., Glerum, A., Stathopoulos, A., Bierlaire, M. & Longo, G. (2017). Mitigating the impact of errors in travel time reporting on mode choice modelling. *Journal of Transport Geography* **62**, 236-246.

Villas-Boas, J.M. (2007). A note on limited versus full information estimation in non-linear models. Discussion Paper, University of California, Berkeley.

Villas-Boas, J.M. & Winer, R.S. (1999). Endogeneity in brand choice models. *Management Science* **45**, 1324–1338.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54**, 426-482.

Walker, J.L. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. Ph.D. Thesis, Massachusetts Institute of Technology.

Walker, J.L. & Ben-Akiva, M.E. (2002). Generalized random utility model. *Mathematical Social Sciences* **43**, 303-343.

Washington, S., Karlaftis, M.G., Mannering, F. & Anastasopoulos, P. (2020). *Statistical and Econometric Methods for Transportation Data Analysis.* 3rd ed. CRC Press, New York.

Wen, C.H. & Chen, P.H. (2017). Passenger booking timing for low-cost airlines: a continuous logit approch. *Journal of Air Transport Management* **64**, 91-99.

Williams, H.C.W.L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Enviroment and Planning Part A: Economy and Space* **9**, 285-344.

Williams, H. C.W.L. & Ortúzar, J. de D. (1982). Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research Part B: Methodological* **16**, 167-219.

Wooldridge, J. (2010). *Econometric Analysis of Cross-Section and Panel Data*. 2nd ed. The MIT Press, Cambridge, Mass.

Wooldridge, J. (2015). Control function methods in applied econometrics. *Journal of Human Resources* **50**, 420-445.

Yáñez, M.F., Raveau, S. & Ortúzar, J. de D. (2010). Inclusion of latent variables in mixed logit models: modelling and forecasting. *Transportation Research Part A: Policy and Practice* **44**, 744–753.

Yatchew, A. & Griliches, Z. (1985). Specification errors in probit models. *The Review of Economics and Statistics* **67,** 134-139.

Zivot, E., Startz, R. & Nelson, C.R. (1998). Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* **39**, 1119-1246.

Zwerina, K., Huber, J. & Kuhfeld, W. (2005). *A General Method for Constructing Efficient Choice Designs.* SAS Institute, Ludwigshafen.