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

Thomas E. Guerrero[1a], C. Angelo Guevara[bc], Juan de Dios Ortúzar[cd], Elisabetta Cherchi[e]

a. Department of Civil Engineering Universidad Francisco de Paula Santander Ocaña, Sede el Algodonal Ocaña, Colombia
b. Department of Civil Engineering, Universidad de Chile
c. Instituto Sistemas Complejos de Ingeniería (ISCI)
d. Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile
e. School of Engineering Future Mobility Group, Newcastle University, Newcastle

ABSTRACT

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 paper, 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.

Keywords: Endogeneity, Linear models, Discrete choice models, Control function, Weak instruments

1. Introduction and Motivation

Endogeneity arises when there is a correlation between the observed explanatory variables and the error term of a model and may cause inconsistent parameter estimation. Endogeneity is considered an unavoidable problem, and there are several reasons behind it, such as omitted attributes, measurement or specification errors, simultaneous determination and/or self-selection (Guevara, 2015). If a model is not adequately corrected for the effect of endogeneity, estimates will be inconsistent, and the model may produce unreliable forecasts and conclusions, leading to potentially wrong decision making (Guevara and Ben-Akiva, 2006).

[1] teguerrerob@ufpso.edu.co (Corresponding Author)

The correction of endogeneity has been extensively studied in linear models (Wooldridge, 2010), but it has been less so in the case of Discrete Choice Models (DCM). DCM are popular in practice to model choices between discrete alternatives (Train, 2009). For example, DCM play a fundamental role in short, and long-term transport planning and policy formulation (Ortúzar and Willumsen, 2011) and are often used to model the third stage of the classic transport model (i.e., in the mode choice model). DCM are very susceptible to endogeneity problems in transport mode choice because it is common to have omitted attributes, measurement or specification errors, simultaneous determination and/or self-selection (Guerrero *et al*., 2020). Notwithstanding, DCM are often used in many other research areas, such as safety (Rizzi and Ortúzar, 2006), environmental studies (Hess and Beharry-Borg, 2012), urbanism (Torres *et al*., 2013) and residential location (Hurtubia and Bierlaire, 2014), so addressing this issue may have widespread implications.

Some approaches have been proposed in the literature to correct for endogeneity in DCM. The most accredited methods are Control Function - CF (Heckman, 1978), Latent Variables - LV (Walker and Ben-Akiva, 2002), Maximum Likelihood (Park and Gupta, 2009; Train, 2009), Berry, Levinsohn and Pakes - BLP - model (Berry *et al*., 1995), Proxies (Guevara, 2015), and Multiple Indicator Solutions - MIS (Guevara and Polanco, 2016). In particular, the CF method presents advantages in many cases, given its easy application, low consumption of computational resources and its ability for correcting endogeneity at the individual level (Guevara, 2015). In practice, the CF approach has been widely used to correct for endogeneity in DCM, among others, by Guerrero *et al*. (2020), Lurkin *et al*. (2017), Mumbower *et al*. (2014) and Wen and Chen (2017), to obtain consistent estimators of the model parameters.

However, like most methods to correct for endogeneity, CF usage involves the challenge of obtaining proper *instruments* (also called instrumental variables – IV; Bresnahan, 1997; Guevara, 2015). The problem is that the instruments must fulfil two seemingly conflicting requirements: (*i*) be correlated with the endogenous variable (relevance condition), and (*ii*) be independent of the DCM error term (exogeneity condition). The *exogeneity condition* for DCM has been recently studied by Guevara (2018) using over-identification tests. In turn, the *relevance condition* has only been incipiently analysed in this context, a gap we contribute to bridge in this article.

The relevance condition is studied in terms of the strength or weakness of the instruments. An instrument is said to be strong when it is sufficiently correlated with the endogenous variable so that it allows estimating consistent parameters. On the other hand, a weak instrument will have poor/slight correlation with the endogenous variable, an undesirable situation that may lead to inconsistent estimates as well (Stock and Yogo, 2005). Correcting endogenous models with weak instruments may lead to models with even worse performance than uncorrected ones (Staiger and Stock, 1997).

The challenge in assessing the relevance condition lies in determining *critical values* (CV), for feasible statistics that can be calculated from available data that allow judging whether an instrument is strong or weak, under certain criteria. Although identifying of weak instruments for correcting endogeneity has been profusely studied in linear models (Stock and Yogo, 2005), it needs to be studied in more depth for DCM. This aspect is relevant because DCM play a vital role in modelling demand and many different subject areas.

This paper 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). As a first step, we validated our empirical approach in linear models and compared them with critical values achieved by Stock and Yogo (2005) and Skeels and Windmeijer (2018) from analytic derivations. Then we tabulate the CV that allows determining if an instrument is weak or not, in the case of a DCM for a single endogenous regressor, varying the number of instruments (k_z) used, and how much RB the modeller considers tolerable. In our article, we corroborate Skeels and Windmeijer (2018) results for two instruments and 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 paper 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) and Skeels and Windmeijer (2018). Then, the third section describes the alternative Monte Carlo method proposed to test for weak instruments and its application to linear models, which serves as a validation, with the addition of the single instrument case. The fourth section presents the application of the proposed Monte Carlo method to test for weak instruments extended and adapted to DCM. Finally, section 5 presents our main conclusions.

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 in detail state of the art to detect weak instrument in linear models.

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. Mainly, this 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 if it was not corrected. 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 the instruments to correct for endogeneity. Initially, R^2 was considered a useful measure of the relevance condition for univariate models, but it was warned that it could be misleading when there are multiple endogenous variables. Staiger and Stock (1997) were the first to define a "*rule of thumb*" coming from an asymptotic distribution resulting from weak instrument problem. This rule established that an instrument is weak when the first-stage F-statistic is less than ten. On the other hand, findings of Zivot *et al.* (1998) recommended to check the performance of the first-stage regression and then making an inference based on 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 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 the 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, later Skeels and Windmeijer (2018) 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 it will be shown 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.

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

Now we describe in more detail the framework proposed for linear models by Stock and Yogo (2005), and improved by Skeels and Windmeijer (2018), and report their findings, which we will use as a benchmark for validation, in more depth. 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 3.

Consider the linear model shown in (1) and the reduced form for the explanatory variable x in (2):

$$y = \theta x + \epsilon \qquad (1)$$

$$x = \pi z + \Delta \qquad (2)$$

where y, Δ 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 $E(\epsilon|\Delta) \neq 0$; therefore, x is correlated with ϵ in (1) because Δ and ϵ have some level of correlation (ρ). Note also that z is correlated with x through (2), but not with ϵ in (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) propose two quantitative criteria to detect weak instruments. Both criteria are derived for the estimator of one or more endogenous regressors. The first is based on the maximum estimator bias, and it is named as *relative bias* (RB), whereas the second is related to the maximum Wald test's *size distortion* (SD). For our research, 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 (3):

$$RB = \left|\frac{E[\hat{\theta}_{TSLS}]-\theta}{E[\hat{\theta}_{OLS}]-\theta}\right| \quad (3)$$

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 the instruments were equal to zero.

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

$$F = \frac{(SSRR-SSRU)}{SSRU}\frac{(N-K)}{k_z} \xrightarrow{d} \chi^2_{k_z}\ , \quad (4)$$

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 1. We reproduce these values in detail because we will use them to contrast our results on weak instruments for DCM.

Table 1 CV of first-stage 95% F-statistic to detect weak instruments for a single endogenous regressor in linear models

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

Various things can be noted in Table 1. The first is that the critical values depend on k_z and RB, and that they grow with the former, and decrease with the latter. 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. Also, note that Stock and Yogo do not show CV for RB of 0.01, 0.15 and 0.25; whereas these are shown by Skeels and Windmeijer. Besides, the CV of Stock and Yogo start at k_z=3, whereas those of Skeels and Windmeijer start from k_z=2.

3. An alternative 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 the Monte Carlo simulation to obtain an empirical distribution of the first stage's F-statistics of the CF method for the desired RB. We then retrieve the percentile 95th, which we considered is 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.

We begin validating the Monte Carlo approach by contrasting the results attained for the linear model, with those that have been obtained from analytical methods in recent literature. Afterwards, we apply the proposed method to a binary logit model, results that can be generalized as will be seen later. Besides, as explained before, in both cases we study the single instrument problem.

The idea behind this process is to construct an experimental setting to validate the methodological approach in the case of linear models and then to extend it to the case of DCM. For this, we will consider the linear model shown in (5) and the reduced form in (6) for an explanatory variable t, which will be regarded as the only endogenous variable in this example. The aim of this model is to emulate the analytic process carried on by Stock and Yogo (2005) and then improved by Skeels and Windmeijer (2018).

$$y = \theta_0 + \theta_c c + \theta_t t + \epsilon \tag{5}$$

$$t = \pi_0 + \pi_{z1} z_1 + \pi_{z2} z_2 + \Delta \tag{6}$$

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 (6) for the explanatory variable t is explained by the intercept term (π_0), a couple of instruments z_1 and z_2, and Δ (error term). For illustrative purposes, we will consider only two instruments because this is the minimum number of instruments 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 it will be shown later, the findings achieved for this validation process were estimated up to fifteen instruments, which seems to be a proper maximum in practical terms.

For the simulation, the data were generated for the maximum level of correlation (ρ=1) between ϵ and Δ; in this way, t and ϵ are correlated in (5) and, therefore, endogeneity arises. The variable y was constructed as a function of c, t and ϵ, with coefficients $\theta_0 = \theta_t = \theta_c = 2$. Δ, 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 due only to wanting to give less variability in the instruments than in the other variables. This fact should not affect the results of our simulation.

The methodological approach for the validation stage is based on the estimation of two linear models: an endogenous and a corrected one. This is because we need to estimate the RB shown in (3). We use the superscript *END* and *CF* to denote if the vector of parameter estimates comes from the endogenous or the corrected model, respectively. The endogenous model is estimated by OLS following the functional form shown in (7), without correcting for endogeneity:

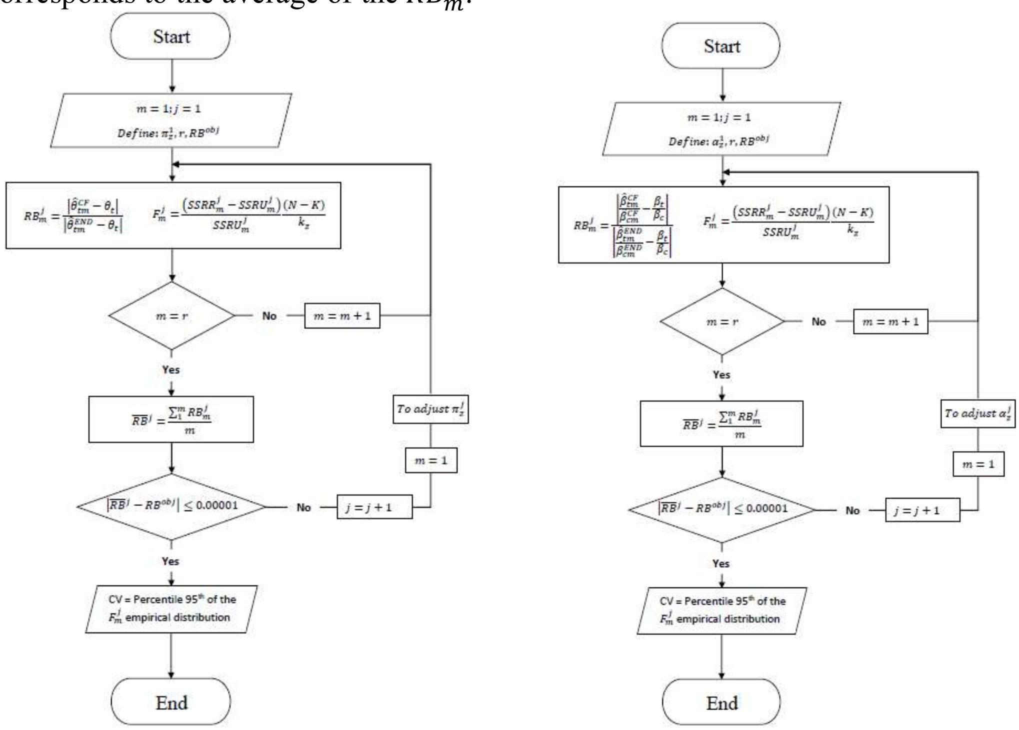
$$y = \hat{\theta}_0^{END} + \hat{\theta}_c^{END} c + \hat{\theta}_t^{END} t + \epsilon \quad (7)$$

We then use the two-stage CF approach (Wooldridge, 2010) to obtain the estimated parameters of the corrected model. In the first stage shown in (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). Then, in the second stage shown in (9), we estimate the linear model considering the residuals ($\hat{\varphi}$) as explanatory variables:

$$t = \phi_0 + \hat{\phi}_{z1} z_1 + \hat{\phi}_{z2} z_2 + \hat{\phi}_c c + \varphi \xrightarrow{OLS} \hat{\varphi} = t - \hat{t} \quad (8)$$

$$y = \hat{\theta}_0^{CF} + \hat{\theta}_c^{CF} c + \hat{\theta}_t^{CF} t + \hat{\theta}_{\hat{\varphi}}^{CF} \hat{\varphi} + \epsilon \quad (9)$$

The estimation of the parameters of the endogenous and corrected models is done through an iterative process, modifying the power of the instruments by adjusting π_{z1} and π_{z2} in (6) until reaching the desired RB (RB^{obj}). The iterative process flowchart is shown in Figure 1(a). The subscript j is used to highlight that π_z in (6) is changed j times until it reaches the desired $\overline{RB}^j$, which corresponds to the average of the RB_m^j.



(a) (b)

Figure 1 Iterative process flowchart to reach the $\overline{RB}^j$ desired in (a) linear and (b) DC models

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 process is repeated *m* times for a set of fixed instrument vectors and a given value of π_z^j, which is 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). Following this approach, for a given set of instruments and a desired RB, we obtain an empirical distribution of F_m^j, from the m repetitions, enabling us to compute the percentile 95th. This value will correspond 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 1(a) is given by (3).

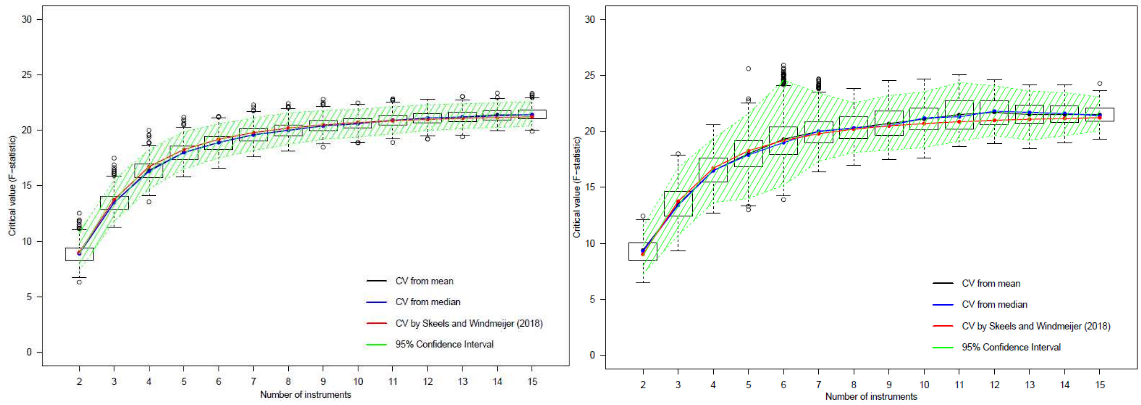
We used the iterative process shown in Figure 1(a) for estimating the CV and then compared our values with those put forward by Stock and Yogo (2005), and Skeels and Windmeijer (2018) for linear models using analytical methods. Our CV were reached for a sample size of 10.000 individuals. The number of repetitions can vary depending on the precision wanted. We considered precision of 0.05 or less on the CV was enough for our results. This fact allowed us to report until one decimal for the CV, for comparison purposes. To achieve this level of precision, we had to perform up to 78,000 simulations, which were subsampled in sets of 100 repetitions. In this way, we obtained 780 values to estimate the mean, median and 95% confidence interval (CI) for the CV, for a single endogenous regressor in a linear model. These values are summarized in Table 2.

Table 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 proposed Monte Carlo Method

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

The estimates show, first, that the means and the medians in linear models are almost identical, and that the 95% CI have widths that vary between 0.9 and 3.72 points. Comparing these simulated results with the findings of Stock and Yogo (2005) and Skeels and Windmeijer (2018), reported in Table 1, we note, first, that in both cases, the CV are larger as the desired RB decreases and the impact of the number of instruments (k_z) also has the same form. Besides, the results of the mean and the median in Table 2 are virtually the same as the values reported in Table 1. Formally, we cannot reject the null hypothesis that the means of the CV obtained with our Monte Carlo 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 3(a) shows boxplots for CV and confidence intervals as a function of the number of instruments, for RB = 0.05 together with the respective values reported by Skeels and Windmeijer (2018). The black and blue dots depict the mean and the median (50th percentile) of all repetitions of the experiment in each boxplot. The abscissa shows the number of instruments k_z, whereas the ordinate corresponds to the critical value obtained in the validation process. As stated, the CV obtained with the Monte Carlo Method are almost indistinguishable from those of Skeels and Windmeijer (2018). The minor differences can be attributed to sampling and simulation errors and the difficulties 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.



(a) (b)

Figure 3 Boxplot for CV and CI by number of instruments and RB 0.05 for (a) linear and (n) DC models

As discussed before, Skeels and Windmeijer (2018) were able to determine the CV for the case of 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 two instruments case (Stock and Yogo, 2005; Angrist and Pischke, 2009). Notwithstanding, our results with the Monte Carlo approach are fully concordant with those attained by Skeels and Windmeijer (2018) using asymptotic theory.

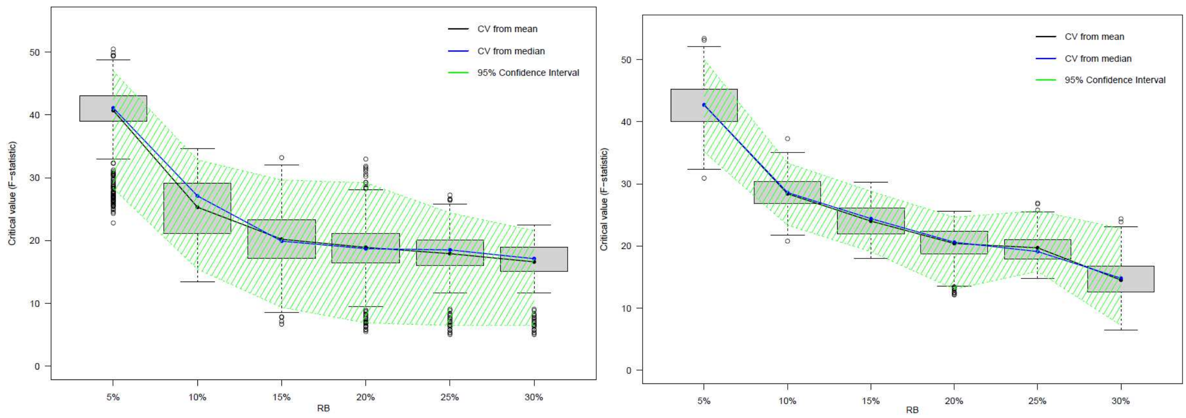
Beyond the k_z=2 case, a more relevant situation for practice would be to provide recommendations for the single instrument case (namely k_z=1), since in most practical applications finding even a single instrument is extremely hard. In this sense, it would be interesting to know if the *rule of thumb* (F-statistic is less than ten), or if values "projected" from the results available for more instruments, would work. Looking at Tables 1, 2 and Figure 3(a), 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 k_z=1 and found CV that were significantly larger than those suggested by a simple extrapolation of the values in Table 1. As a preliminary validation of this seemingly weird result, they report the results of a Monte Carlo analysis for RB which values are 0.01, 0.05, 0.10 and 0.20.

Our Monte Carlo method has no limits in analysing the case of k_z=1 and a single endogenous variable. The mean, median and confidence intervals for critical values in this case from our simulation, are reported in Table 3.

Table 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

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]

The boxplots displayed in Figure 4(a) are drawn for the same simulation conditions described above for $k_z \geq 2$. As can be seen, the critical values reached with our empirical approach are far from those that would be suggested by a naive extension of the results for a larger number of instruments. They imply that the rule of thumb of F>10 is far from being enough at least for this case, and the CV to consider should be three times larger than that. Interestingly, our results in this regard coincide with those reported by Skeels and Windmeijer (2018) in their Table A1 of Appendix D, confirming their (and our) results.



(a) (b)

Figure 4 Boxplot for CV when k_z =1 and RB=0.05 in (a) linear and (b) DC models

4. Application of the proposed Monte Carlo method to test for weak instruments in DCM

To extend and adapt our empirical methodology to the case of DCM, we started by using 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 (10).

$$U_{in} = ASC_i + \beta_c c_{in} + \beta_t t_{in} + \varepsilon_{in} \quad (10)$$

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 $\beta_c = 2$. Given that the error term ε_{in} in (10) distributes Normal, the choice model is formally a Probit model. The model was estimated as a binary Logit to simplify estimation. This assumption can be made given that Lee (1982), Ruud (1983) and Cramer (2007) state that it is not a severe problem because it does not compromise the possibility to obtain consistent parameters up to a scale.

Afterwards, 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 Logit for simplicity, as will be shown below, it can be extended to multinomial DCM as will be seen later.

We built our simulation experiment for a single endogenous variable; in particular, we assumed that t_{in} in (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 (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.

$$t_{in} = \alpha_0 + \alpha_{z1} z_{1in} + \alpha_{z2} z_{2in} + \xi_{in} \quad (11)$$

Endogeneity arises in our simulation because we generated correlation between the terms ε_{in} and ξ_{in}, $[corr(\varepsilon_{in}, \xi_{in}) = \rho \neq 0]$. Therefore, t_{in} is correlated with ε_{in} in the utility function (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 $\alpha_0 = 1$.

Following the same methodological approach used in the case of linear models above, we proposed the estimation of 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 (12), which is not corrected for endogeneity.

$$U_{in} = \widehat{ASC}_i^{END} + \hat{\beta}_c^{END} c_{in} + \hat{\beta}_t^{END} t_{in} + \varepsilon_{in} \quad (12)$$

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

The first stage, shown in (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 (14). This two-stage procedure does not compromise estimator's consistency, which is guaranteed by the Slutsky theorem (Ben-Akiva and Lerman, 1985).

$$t_{in} = \hat{\gamma}_t + \hat{\gamma}_{z1} z_{1in} + \hat{\gamma}_{z2} z_{2in} + \hat{\gamma}_c c_{in} + \delta_{in} \xrightarrow{OLS} \hat{\delta}_{in} = t_{in} - \hat{t}_{in} \quad (13)$$

$$U_{in} = \widehat{ASC}_i + \hat{\beta}_c^{CF} c_{in} + \hat{\beta}_t^{CF} t_{in} + \beta_{\hat{\delta}} \hat{\delta}_{in} + \tilde{\varepsilon}_{in} \quad (14)$$

The iterative process flowchart to determine critical values in the case of DCM is displayed in Figure 1(b). The iterative process is the same for linear models shown above. However, as correcting for endogeneity in DCM implies a change of the scale in the estimators (Guevara and

Ben-Akiva, 2012), the ratio among parameters (i.e., β_t/β_c) must be checked instead of the parameters directly. Therefore, the proper equation in Figure 1(b) to calculate the RB_m^j for DCM suffers a change regarding the one shown in Figure 1(a).

Here, we continue considering a precision of 0.05 or less for the CV was enough for our results in DCM, for being able to report until one decimal for the CV. 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.

Table 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 proposed Monte Carlo Method

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

As can be seen, the median and mean from the simulation tends to be the same or very similar. Notwithstanding, we recommended 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 proportions of extreme observations known as 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 a 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, 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 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 3(b) 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. 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 that the dots representing the mean and median (black and blue lines, respectively) are contained within the CI. Thereby, it cannot be rejected the null hypothesis that the mean and median are the statistically same compared to the CV reached by Skeels and Windmeijer (2018) for the linear models.

Figure 3(b) shows the presence of outliers in the empirical distribution of the CV. This fact can be due to these values are reached as part of a random process from the Monte Carlo simulation. Besides, the correction process for endogeneity with weak instruments involves difficulties, especially in DCM. Many of these outliers can be seen when the k_z is six and seven. For these special cases, we recommended using the median instead of the mean.

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 simply impossible (Guevara, 2015). In practical terms, it would be useful to know the CV from 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. Mean, median and CI for the CV achieved from our empirical approach for a single endogenous regressor in DCM are shown in Figure 4(b) and reported in Table 5.

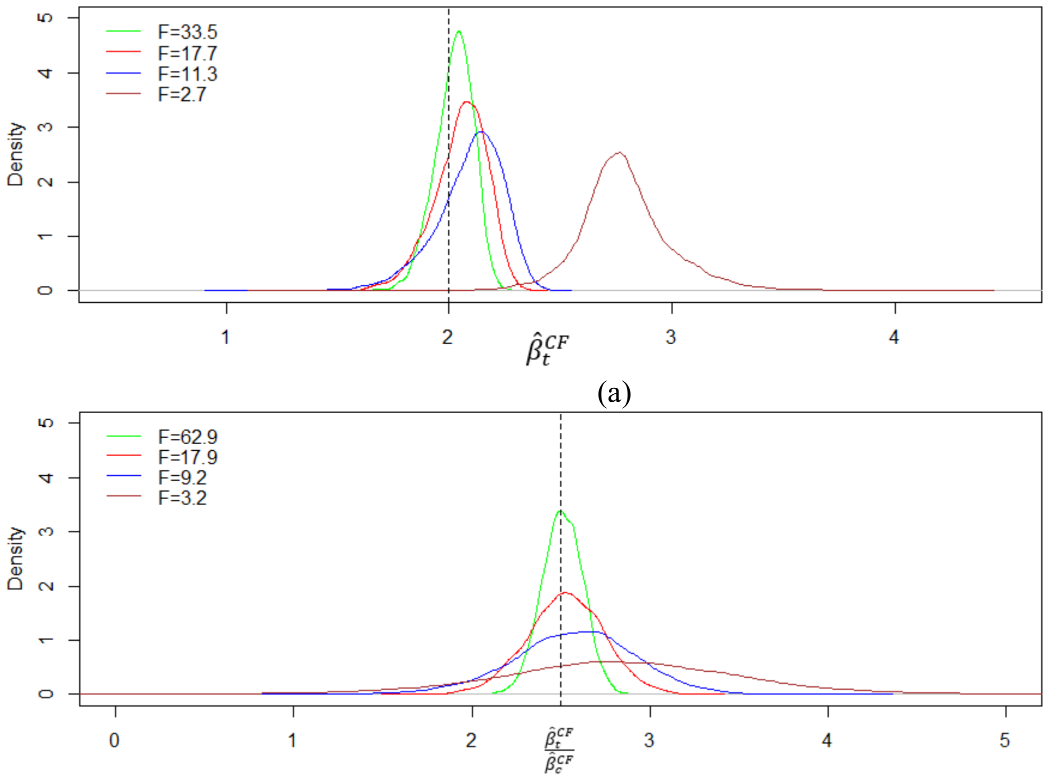
Table 5 Mean, median and CI for the CV of first-stage 95% F-statistic to detect weak instruments for a single endogenous regressor in DCM with proposed Monte Carlo Method for k_z=1

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

Interestingly, the CV estimated for the DCM has the same order of magnitude of those achieved for linear models. Skeels and Windmeijer (2018) recommend that their findings for k_z =1 in linear models are useful for RB's absolute values less than 0.20. This suggestion is because the TSLS estimator coming from their simulation shows a standard deviation large. In our case, we achieved to reach CV from our methodological approach for RB to up 0.20, even 0.25 and 0.30. In any case, our results for k_z =1 in DCM are in line with those represented by Skeels and Windmeijer (2018) in Table A1 of Appendix D.

The experiments presented in this section consisted of binary logit models, but due to the large computational burden of the proposed approach, it was almost prohibitive. We consider that there is no reason for not extending our methodological approach from the binary to the multinomial DCMs. For illustrative purposes, we explored the impact of extending this analysis to a DCM with five alternatives. We kept the same simulation conditions used with the binary logit model. We analyzed the case k_z=6 and RB=0.10. The findings shown in Table 4 for the binary logit model, we knew that the mean is 11.0, and the median is 10.9. The estimates of the mean and median for the case of the MNL of 5 alternatives, k_z=6 and RB=0.10 are 11.04 and 10.93, respectively. We applied the t-statistics (Ortúzar and Willumsen, 2011, pages 341-342) to compare if the means and medians from binary and MNL models are the statistically same. This way, the H_0 is accepted and determine that they are not statistically different from each other. Therefore, we can conclude that our findings can be extended to the case of logit models with multiple alternatives.

On the other hand, we also investigated the effects of the weak instruments on the estimators. For this, we did simulations for the case of RB=0.05 and k_z=5. The simulations' idea was to vary the values of α_z that allowed to reach the mean and median reported in Table 4 (18.0 and 17.9, respectively). In this way, we obtained empirical estimator distributions with and without weak instruments, as can be seen in Figure 8 (b).



(a)

(b)

Figure 8 Effect of the weak instruments on the estimator distribution in (a) linear and (b) DC models

As shown in Figure 8 (b), the abscissa corresponds to the ratio $\hat{\beta}_t^{CF}/\hat{\beta}_c^{CF}$, whose real value is 2.5 (β_t/β_c), and it is shown as a black dashed line. On the other hand, the ordinate indicates the estimator density. 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 it has relatively symmetrical tails. However, as α_z decreases (and therefore F-statistic), the correction's quality worsens, the variance increases, and the estimator's sample distribution deforms. The red line corresponds to the estimator's sample distribution for the critical value when k_z=5 and the RB=0.05. Under these conditions, the estimator's sample distribution still retains a defined "*bell*" shape of a Normal distribution. It is a relevant finding because this behaviour does not happen for the linear models, as shown in Figure 8(a). Apparently, for linear models affected by weak instruments, the sample distribution of the estimator does not deform. It seems that by decreasing the F-statistic, the variance increases slightly; however, the correction is not better. We think that these differences between linear models and DCM could be due to the scale's change in the estimators that involve the DCM (Guevara and Ben-Akiva, 2012).

5. Conclusions

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 it cannot be neglected the adverse effects of weak instruments in modelling. As in linear models, also in DCM, weak instruments affect the estimation of consistent parameters.

This paper 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 result found by Guevara and Navarro (2015) based on the findings of Stock and Yogo (2005) and improved by Skeels and Windmeijer (2018) in the case of the linear models. Our results are supported by the validation of our simulation approach for linear models and then applied to DCM. The CV for DCM are very similar to those found for linear models. Formally, we cannot reject the hypothesis that they are equal. For more details, the reader can refer to Table 2 and 4. Our research's additional contribution is finding the determination of the CV for k_z=1 for a single endogenous variable in the linear model and DCM (See Tables 3 and 5). Our findings are in line with those reported by Skeels and Windmeijer (2018) in Table A1 of Appendix D. These findings are relevant, overall because in practice, there are difficulties often encountered in finding a sufficient number of instruments that fulfil the relevance condition. It is a significant and useful finding.

Second, we show evidence of the adverse effects of weak instruments on the DCM estimator's sample distribution. Some anomalies were found, such as quality loss of the correction for endogeneity, variance increase, and estimator's sample distribution deformation. This fact is relevant because of the effect on statistical inference (e.g. estimators, confidence intervals and tests of hypotheses). The results in linear models seem to be slightly different but not worse. To the best of our knowledge, these findings have not been discussed in the literature before. The CV depend strongly on the RB that the modeller is willing to tolerate and the k_z. Our results also show a loss of Power as more RB is accepted.

The use of Monte Carlo experiments in our simulation proposal represents the third contribution. We determined that it works properly because the simulation proposal was validated in linear models and then extended to DCM. The validation stage showed that the CV for linear models from our approach presents a slight difference compared to those determined analytically and from asymptotic distributions (Stock and Yogo, 2005; Skeels and Windmeijer, 2018). Despite this, the hypothesis that the CV coming from the validation process with our simulation approach are the same regarding the CV reached by Skeels and Windmeijer (2018) for linear models cannot be rejected. We consider both approaches are proper.

Our findings are valuables; however, this research has some limitations. Possible research extensions are also proposed. The CV for DCM coming from Monte Carlo simulations, contrary to the analytic derivation made for linear models by Stock and Yogo (2005) and Skeels and Windmeijer (2018). We do not consider the size distortions of the Wald statistic. This criterion was considered by Stock and Yogo (2005). The CV determined are for a single endogenous regressor. Two or more regressor should be considered, which case the Cragg–Donald statistic would be proper instead of the F-statistic. We suggest other types of models and functional forms.

References

Andrews, I., Stock, J. H., and Sun, L. (2019). Weak Instruments in Instrumental Variables Regression: Theory and Practice. *Annual Review of Economics*, **11**, 727-753.

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

Ben-Akiva, M. E., y Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT press.

Bound, J., Jaeger, D. A., and 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.

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

Bresnahan, T.F. (1997). The apple-cinnamon cheerios war: valuing new goods, identifying market power, and economic measurement. In T.F. Bresnahan and R.J. Gordon (eds.), *The Economics of New Goods*, NBER Studies in Income and Wealth Number 58, The University of Chicago Press. Chicago.

Cameron, A. C., and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

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

Cragg, J. G., and Donald S. G. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory*, **9**, 222–40.

Cramer, J. (2007). Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances. *Oxford Bulletin of Economics and Statistics*, **69**, 545-555.

Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, vol. 1. Cambridge University Press, Cambridge.

Dinner, I. M., Heerde Van, H. J., and Neslin, S. A. (2014). Driving online and offline sales: The cross-channel effects of traditional, online display, and paid search advertising. *Journal of marketing research*, **51**, 527-545.

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

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

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.

Guevara, C.A. (2015). Critical assessment of five methods to correct for endogeneity in discrete-choice models. *Transportation Research Part A*, **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. and Ben-Akiva, M. (2006). Endogeneity in residential location choice models. *Transportation Research Record*, **1977**, 60–66.
Guevara, C.A., and 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. and Navarro, P. (2015) Detection of weak instruments when correcting for endogeneity in binary logit models. *Proceedings 14th International Conference on Travel Behaviour Research (IATBR)*, Windsor, UK, July 19-23.
Guevara, C.A. and 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.
Heckman, J.J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, **46**, 931–959.
Hess, S., and 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.
Hurtubia, R., and 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.
Kinal, T. W. (1980). The existence of moments of k-class estimators. *Econometrica*, **49**, 241–249.
Lee, L. (1982). Specification Error in Multinomial Logit Models. *Journal of Econometrics*, **20**, 197–209.
Magnusson, L. M. (2007). Weak instruments robust tests for limited dependent variable models. *Working Paper*, Brown University (RI).
MathWorks. (2016). *MATLAB and Statistics Toolbox Release 2016b*. Natick: The MathWorks Inc.
Mumbower, S., Garrow, L.A. and 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.
Nelson, C. R. and Startz, R. (1990a). Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator. *Econometrica*, **58**, 967–976.
Nelson, C. R. and 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.
Lurkin, V., Garrow, L.A., Higgins, M.J., Newman, J.P. and 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.
Ortúzar, J. de D. and Willumsen, L.G. (2011). *Modelling Transport* (4th ed). Chichester: John Wiley & Sons.
Park, S. and Gupta, S. (2009). Simulated maximum likelihood estimator for the random coefficients logit model using aggregate data. *Journal of Marketing Research*, **46**, 531-542.
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, Vienna. Retrieved December 2010 from http://www.Rproject.org.
Rizzi, L., and Ortúzar, J. de D. (2006). Road safety valuation under a stated choice framework. *Journal of Transport Economics and Policy*, **40**, 69-94.
Rothenberg, T. J. (1984). Approximating the distributions of econometric estimators and test statistics. *Handbook of econometrics*, 2, 881–935. Amsterdam: North Holland.