Documento de Trabajo 2012-01
Facultad de Economía y Empresa
Universidad de Zaragoza
Depósito Legal Z-1411-2010. ISSN 2171-6668

SELECTION CRITERIA FOR OVERLAPPING BINARY MODELS

M.T. Aparicio and I. Villanúa

Department of Economic Analysis, Faculty of Economics, University of Zaragoza, Gran Vía 2-4, 50005, Zaragoza, Spain.

Abstract

The binary model selection procedures are the purpose of this paper. It has been very often studied in the case of linear and nested competing models, but not too much in the framework of non linear and non nested models. Using the classification of Vuong (1989) in nested, overlapping and strictly non-nested models, we focus on the overlapping models. The special situation that the competing models don't include the true model is studied together with the usual case of the true model included in the compared models. We carry out the analysis both theoretically and through a Monte Carlo experiment.

JEL Classification: C15, C25, C44

Key Words: logit and probit, selection criteria, overlapping models, true model, convergence, asymptotic behaviour.

Corresponding author: aparicio@unizar.es

Acknowledgement: The authors wish to thank the financial support provided by the Spanish Department of Science and Technology, Project ECO 2009-07936 and from the Aragon's Government, Project S21.

SELECTION CRITERIA FOR OVERLAPPING BINARY MODELS

1. INTRODUCTION

This work focuses on the analysis of model selection criteria within the framework of Binary Choice Models (BCM). In a general context, Vuong (1989) distinguishes between nested, overlapping and non-nested models. Following this classification, we are interested in the overlapping BCM, and we can say that two binary models are overlapping if both have the same functional form (both probit or both logit) with some common explanatory variables and other specific variables. If we compare overlapping models, it can happen that they were equivalent or one of them better than the other (non-equivalent).

Although the Hypothesis Testing procedures (HTP) are very used to discriminate between models, they only allow us to choice between pairs of models. Against them, the selection criteria allow us to select the best model among a quite large set of them. This is an important advantage in empirical econometric works. The latter approach allows researchers to express their objectives in the form of a loss function, or by using discrepancy concept[1]. For non-linear regression models, which is our framework, the procedures developed by Vuong (1989), Pesaran and Pesaran (1993) and Santos Silva (2001) belong to the first category (HTP). The second category involves the well-known AIC (Akaike, 1973) and SBIC (Schwarz, 1978), where the discrepancy is obtained from the Kullback Leibler distance. Additionally, the use of the Mean Square Error of Prediction as a discrepancy allows us to derive another criterion, denoted as C_2 (see Aparicio and Villanúa (2007)).

Many works have studied the behaviour of selection procedures in linear regression models. However, this subject has been less analysed in a non-linear regression context, and the nested framework is nearly always assumed (Kim and Gavanaugh (2005) and Van der Hoeven (2005)). Specific references for discrete choice models are Aparicio and Villanúa (2007) for nested models and Monfardini (2003) for non-nested models.

In this paper, the competing models we select from are overlapping models. The purpose is to investigate the discrimination power of some model selection criteria assuming two situations: i) At least one of the models is correctly specified; ii) Neither of the models are correctly specified. According Bierens (1994), a well- specified model can include irrelevant variables together the set of regressors of the Data Generating Process (DGP). In our opinion,

[1] As Linhart and Zucchini (1986) and Lavergne (1998) established, the discrepancy concept is a particular case of loss function

the second situation is the most interesting in practice but the least studied in the literature. Given that in this case none model is well specified, we can't consider consistency as the condition making a given selection criterion adequate. The requirement we propose is that the criterion selects the closest model to the DGP. This same condition will be used when one of the models was correctly specified, which is identical to demand consistency.

The article is organised as follows. In Section 2 we establish the general context, in Section 3 we study the theoretical behaviour of the criteria. Section 4 presents a Monte Carlo experiment and conclusions are presented in Section 5.

2. THE THEORETICAL FRAMEWORK

Our framework is made up of a pair of overlapping models which we define in general terms as follows:

M1: $p_i = F(a_i'\gamma)$

M2: $p_i = F(b_i'\delta)$ (1)

where $F(\cdot)$ is the cumulative distribution function (c.d.f) , which can be normal or logistic, leading to the probit model or the logit model, respectively. The two competing models have the same c.d.f, that is to say, both are probit, or both are logit; a_i' and b_i' are the $1\times k_1$ and $1\times k_2$ explanatory variables vectors of M1 and M2, for the i-observation, with i=1,…,N; γ and δ are the corresponding parameter vectors. Given the definition of overlapping models, it is satisfied $a' \not\subset b'$ and $b' \not\subset a'$, and both vector have some common variables.

The DPG is denoted as:

M0: $p_i = F(x_i'\beta^0)$ (2)

where the explanatory variables vector x_i' is $1\times k_0$, and β^0 is the true parameter vector $k_0 \times 1$.

For the purpose of ordering the competing models, the relationship of each of them with the true model (DGP) must be established. To achieve this, we use the Kullback-Leibler distance (KLD) to the DGP, which is written as follows:

$$\text{KLD (M0, M}j\text{)} = E_0 \ln\left[\frac{f_0}{f_j}\right] \qquad (j = 1,2) \qquad (3)$$

being f_0 the density function of the DGP and f_j corresponding to model Mj.

The parameter vector that minimises the distance of Mj to M0 is called the pseudo-true parameter vector, and is denoted by γ^* and δ^* for M1 and M2 respectively (see White (1982)).

From (3), we can write:

$$\text{KLD (M0, M1)} - \text{KLD(M0,M2)} = E_0\ [\ln\ f(y \mid b,\ \delta^*)\ ] - E_0\ [\ \ln f(y \mid a,\ \gamma^*)\ ]$$

$$= E_0\ [\ \ell_2^* - \ell_1^*\ ] \qquad (4)$$

If (4) is positive, then M2 is preferred to M1 and the opposite occurs if (4) is negative. If (4) is null, it means that the two specifications are equivalents.

Now, the set of situations to analyse is derived by combining two aspects:

a) The information about the DGP (variables and parameters).
b) The regressors of the competing models.

We assume that the DGP given in (2) is:

$$M0: p_i = F\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right) \qquad (5)$$

and the resulting cases are**:**

<u>Case 1</u>: Neither of the competing models is exactly the DGP. In this context we can consider three alternatives:

Alternative 1.1:

$$\text{M1: } p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + d_i'\gamma_j)$$

$$\text{M2: } p_i = F(\delta_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \delta_3 z_i) \qquad (6)$$

Alternative 1.2:

$$\text{M1: } p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 w_i)$$

$$\text{M2: } p_i = F(\delta_0 + \delta_1 x_{1i} + \delta_2 z_i) \qquad (7)$$

Alternative 1.3:

$$\text{M1: } p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 w_i)$$

$$\text{M2: } p_i = F(\delta_0 + \delta_1 x_{2i} + \delta_2 w_i) \qquad (8)$$

where d_i', z_i and w_i are the additional variables of the competing models. Given that d_i' is a vector and z_i only one variable, it is satisfied that $z_i \not\subset d_i'$.

In alternative 1.1, the DGP is nested in both models M1 and M2. In alternative 1.2 the DGP is nested in one model (M1) only, while in the third alternative the DGP is not nested in any model. According Bierens(1994), in alternative 1.1 both models are well-specified, while in alternative 1.2, only model M1 is correctly specified. In alternative 1.3 neither of the models is correct.

Case 2: One of the competing models is the DGP.

M1: $p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i})$

M2: $p_i = F(\delta_0 + \delta_1 x_{1i} + \delta_2 z_i)$ (9)

This last framework is a particular case of alternative 1.2, given that M1 is well-specified because it is exactly the DGP.

In order to derive a selection criterion, a discrepancy must be defined. If we consider the distance of Kullback Leibler (Linhart and Zucchini (1986)) as a discrepancy, we obtain the well-known AIC and SBIC:

$$\text{AIC (M}j) = -\frac{\hat{\ell}_j}{N} + \frac{k_j}{N} \tag{10}$$

$$\text{SBIC (M}j) = -\frac{\hat{\ell}_j}{N} + \frac{k_j}{2}\frac{\log(N)}{N} \tag{11}$$

where $\hat{\ell}_j$ denotes the log-likelihood of model Mj, evaluated at the corresponding vector of estimates, and k_j is the number of parameters of such model.

Now, if we consider the Mean Square Error of Prediction as a discrepancy, it leads to obtaining the criterion we call C_2, which is written as follows:

$$C_2(\text{M}j) = \frac{SSD_j}{N}\left(1 + \frac{2k_j}{N}\right) \tag{12}$$

where SSD_j is the squared sum of the differences between the binary variable and the estimated probability with model M*j*. The proof of (12) is developed in Appendix 1.

The model having the criterion with the lesser value is chosen, so different criteria can lead to different choice.

As we previously mentioned, our objective is to analyse the theoretical behaviour (in an asymptotic context given that it is the only framework where we know the properties) of the selection criteria. To do that, we must obtain the probabilistic limits *plim*(R(M1)) and *plim*(R(M2)), with R being a given criterion. These probabilistic limits let us know which model the criterion chooses, and compare that choice with what it should have chosen given the true relationship between the competing models (both equivalents, or one better than the other). This is our definition of an adequate criterion: one which selects the model closest to the DGP.

3. THEORETICAL BEHAVIOUR OF THE SELECTION CRITERIA

In order to analyze the asymptotic behaviour of R(M1) – R(M2), we must first specify this difference for each criterion and for every case.

For the AIC criterion, we use (10) to obtain:

$$\mathrm{AIC(M1)} = -\sum_{i=1}^{N}\left[y_i \log F(a_i'\hat{\gamma}) + (1-y_i)\log(1-F(a_i'\hat{\gamma}))\right] + \frac{k_1}{N} = -\frac{\hat{\ell}_1}{N} + \frac{k_1}{N} \tag{13}$$

$$\mathrm{AIC(M2)} = -\sum_{i=1}^{N}\left[y_i \log F(b_i'\hat{\delta}) + (1-y_i)\log(1-F(b_i'\hat{\delta}))\right] + \frac{k_2}{N} = -\frac{\hat{\ell}_2}{N} + \frac{k_2}{N} \tag{14}$$

Similarly, the corresponding expressions for the SBIC criterion are:

$$\begin{aligned}\mathrm{SBIC(M1)} &= -\sum_{i=1}^{N}\left[y_i \log F(a_i'\hat{\gamma}) + (1-y_i)\log(1-F(a_i'\hat{\gamma}))\right] + \frac{k_1}{2}\frac{\log(N)}{N} \\ &= -\frac{\hat{\ell}_1}{N} + \frac{k_1}{2}\frac{\log(N)}{N}\end{aligned} \tag{15}$$

$$\begin{aligned}\mathrm{SBIC(M2)} &= -\sum_{i=1}^{N}\left[y_i \log F(b_i'\hat{\delta}) + (1-y_i)\log(1-F(b_i'\hat{\delta}))\right] + \frac{k_2}{2}\frac{\log(N)}{N} \\ &= -\frac{\hat{\ell}_2}{N} + \frac{k_2}{2}\frac{\log(N)}{N}\end{aligned} \tag{16}$$

From (13) to (16) we can see that AIC criterion differs from SBIC criterion only in the correction factor. As we consider the criteria jointly, we denote them as Information Criteria (IC); we only study AIC and SBIC criteria separately when the correction factor is relevant.

Finally, for the C_2 criterion we obtain:

$$C_2(\mathrm{M1}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - F(a_i'\hat{\gamma})\right)^2 + \frac{2k_1}{N}\sum_{i=1}^{N}\left(y_i - F(a_i'\hat{\gamma})\right)^2 = \frac{SSD_1}{N}\left(1+\frac{2k_1}{N}\right) \tag{17}$$

$$C_2(\mathrm{M2}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - F(b_i'\hat{\delta})\right)^2 + \frac{2k_2}{N}\sum_{i=1}^{N}\left(y_i - F(b_i'\hat{\delta})\right)^2 = \frac{SSD_2}{N}\left(1+\frac{2k_2}{N}\right) \tag{18}$$

We make the following assumption for the cases 1 and 2 established in the previous section :

(A.1) The x_i', a_i' and b_i' vectors are non stochastic. The variables of these vectors have sample means and variances with a finite limit.

Additionally, some of the derivations we will carry out need to take the following lemma into account:

Lemma 1 (Newey and McFadden 1994, p.2156):

"If z_i is i.i.d., $a(z,\ \theta)$ is continuous at θ_0 with probability one, and there is a neighbourhood Γ of θ_0 such that $E\ [\sup_{\theta\in\Gamma}\|a(z,\theta)\|]\ <\ \infty$, then for any $\tilde{\theta}\xrightarrow{p}\theta_0$, $N^{-1}\sum_{i=1}^{N}a(z_i,\tilde{\theta})\xrightarrow{p}E[a(z,\theta_0)]$".

Instead of z_i, our variable is y_i, which is not i.i.d., but heterogeneous (non-identical means and non-identical variances). This fact changes the conclusion of the lemma in the following sense:

$$N^{-1}\sum_{i=1}^{N}a(z_i,\tilde{\theta})\xrightarrow{p}E\left[\frac{1}{N}\sum_{i=1}^{N}a(z_i,\theta_0)\right] \tag{19}$$

The proof of (22) is based on a Law of Large Numbers for heterogeneous variables.[2]

3.1. <u>Alternative 1.1</u>

As we can see in (6), the competing models are:

M1: $p_i = F(\gamma_0+\gamma_1 x_{1i}+\gamma_2 x_{2i}+d_i'\alpha)\ = F\ (a_i'\gamma\ )$

M2: $p_i = F(\delta_0+\delta_1 x_{1i}+\delta_2 x_{2i}+\delta_3 z_i)\ = F\ (b_i'\delta\ )$

in such a way that the DGP is nested in both M1 and M2 models. According to Kohn (1983), the following results of convergence can be deduced:

$$\hat{\gamma}\xrightarrow{p}\gamma^* = \beta^{0+} \tag{20}$$

$$\hat{\delta}\xrightarrow{p}\delta^* = \beta^{0+} \tag{21}$$

where β^{0+} is extended with elements equal to zero in the places corresponding to the variables that are not included in the DGP, that is to say, (β^{0+})' = ($\beta^{0'}$ | 0).

The convergence results (20) and (21) allow us to conclude that both specifications are asymptotically equivalents.

From (4), the following result is satisfied if M1 and M2 are equivalent:

---

[2] This Law can be expressed in the following terms (Davidson 2000, p. 41):

"Let the sequence $\{y_i-\mu_i\}$ be independent with $E\ (y_i-\mu_i) = 0$.

If $E\ |\ y_i-\mu_i\ |^{1+\delta}\le B<\infty\quad\forall i$ with $\delta > 0$, then $\frac{1}{N}\sum_{i=1}^{N}\ (y_i-\mu_i)\xrightarrow{p}0$".

$$E_0\left[\log\frac{f\left(b_i'\delta^*\right)}{f\left(a_i'\gamma^*\right)}\right]=0$$

which can be written as:

$$E_0\left[Y_i\log F\left(b_i'\delta^*\right)+(1-Y_i)\log\left(1-F\left(b_i'\delta^*\right)\right)\right]=E_0\left[Y_i\log F\left(a_i'\gamma^*\right)+(1-Y_i)\log\left(1-F\left(a_i'\gamma^*\right)\right)\right] \tag{22}$$

or equivalently:

$$E_0\left[Y_i\log\frac{F_{1i}}{F_{2i}}+\left(1-Y_i\right)\log\frac{1-F_{1i}}{1-F_{2i}}\right]=0$$

where, $F_{1i}=F\left(a_i'\gamma^*\right)$ and $F_{2i}=F\left(b_i'\delta^*\right)$.

We know that $E_0\left(Y_i\right)=F\left(x_i'\beta^0\right)$ and we call $F_{oi}=F\left(x_i'\beta^0\right)$, so the expression of (22) is:

$$\left[F_{0i}\log\frac{F_{1i}}{F_{2i}}+\left(1-F_{0i}\right)\log\frac{1-F_{1i}}{1-F_{2i}}\right]=0 \tag{23}$$

The following definition establishes what the behaviour of the criteria should be in order to detect the equivalence relationship between the competing models.

DEFINITION 1. Given two models M1 and M2 which are equivalent, a selection criterion (R) is adequate if:

$$plim\ [\ \mathrm{R(M1)}]\ = plim\ [\mathrm{R(M2)}] \tag{24}$$

that is to say, the criterion can select either one.

Next, we analyze the behaviour of each criterion.

- *Behaviour of the information criteria*

We deal jointly with AIC and SBIC, so:

$$\begin{aligned}\mathrm{IC(M1)}&=-\frac{\hat{\ell}_1}{N}+\frac{K_N\left(M1\right)}{N}\\ \mathrm{IC(M2)}&=-\frac{\hat{\ell}_2}{N}+\frac{K_N\left(M2\right)}{N}\end{aligned} \tag{25}$$

where $K_N\left(M_j\right)$ is the correction factor: K_j in AIC, and $K_j\dfrac{\log N}{2}$ for the SBIC criterion.

Given definition 1.1, we need to obtain the probabilistic limit of IC(M1) and IC(M2). If the IC criteria perform adequately, then *plim*(IC(M1) = *plim*(IC(M2). To reach this, we use the expression:

$$\frac{\hat{\ell}_1}{N}=\frac{1}{N}\sum_{i=1}^{N}\left[Y_i\log F\left(a_i'\hat{\gamma}\right)+\left(1-Y_i\right)\log\left(1-F\left(a_i'\hat{\gamma}\right)\right)\right]$$
$$\frac{\hat{\ell}_2}{N}=\frac{1}{N}\sum_{i=1}^{N}\left[Y_i\log F\left(b_i'\hat{\delta}\right)+\left(1-Y_i\right)\log\left(1-F\left(b_i'\hat{\delta}\right)\right)\right] \tag{26}$$

Now, we use Lemma 1 to obtain the following convergence results:

$$\frac{\hat{\ell}_1}{N}\xrightarrow{p}\frac{1}{N}\sum_{i=1}^{N}E_0\left[Y_i\log F\left(a_i'\gamma^*\right)+\left(1-Y_i\right)\log\left(1-F\left(a_i'\gamma^*\right)\right)\right]$$
$$\frac{\hat{\ell}_2}{N}\xrightarrow{p}\frac{1}{N}\sum_{i=1}^{N}E_0\left[Y_i\log F\left(b_i'\delta^*\right)+\left(1-Y_i\right)\log\left(1-F\left(b_i'\delta^*\right)\right)\right] \tag{27}$$

Given that $\frac{K_N\left(M_j\right)}{N}\xrightarrow{p}0$, definition 1 will be satisfied if the two convergences of (27) are the same, in other words, if satisfied:

$$\frac{1}{N}\sum_{i=1}^{N}E_0\left[Y_i\log F_{1i}+\left(1-Y_i\right)\log\left(1-F_{1i}\right)\right]=\frac{1}{N}\sum_{i=1}^{N}E_0\left[Y_i\log F_{2i}+\left(1-Y_i\right)\log\left(1-F_{2i}\right)\right]$$

or similarly:

$$\frac{1}{N}\sum_{i=1}^{N}\left[F_{0i}\log\frac{F_{1i}}{F_{2i}}+\left(1-F_{0i}\right)\log\frac{\left(1-F_{1i}\right)}{\left(1-F_{2i}\right)}\right]=0 \tag{28}$$

This result is always true, because our starting point was that M1 and M2 are equivalent models so expression (23) was satisfied. Now, we can observe that (23) and (28) are identical, and affirm that

$$plim[\mathrm{IC(M1)}]=plim[\mathrm{IC(M1)}] \tag{29}$$

meaning that the Information Criteria AIC and SBIC behave adequately under conditions of alternative 1.1.

Note that the correction factor establishes the convergence rate to achieve the equality (29). From (27), and given (20) and (21), we obtain that $plim\left(-\frac{\hat{\ell}_1}{N}\right)=plim\left(-\frac{\hat{\ell}_2}{N}\right)$. The two expressions of (25) have the same probabilistic limit, but the convergence rate to achieve (29) depends on the limit of $\frac{K_N\left(M_j\right)}{N}$. For the AIC criterion, $K_N\left(M_j\right)$ is $o\left(\log\left(N\right)\right)$, while for the SBIC criterion the same term is $o\left(N\right)$, so $\frac{K_N\left(M_j\right)}{N}$ for AIC converges to zero faster than for SBIC. Then, for finite sample sizes, the SBIC will select the model with fewer parameters (k_j) more times, because $\frac{K_N\left(M_j\right)}{N}$ will converge slower than in the AIC criterion.

If we now focus on only a specific criterion (AIC or SBIC), the convergence rate $\frac{K_N\left(M_j\right)}{N}$ allows us to conclude that the criterion, for finite sample sizes, will choose the model with less k_j, because the correction factor will converge to zero more slowly.

In other words, the IC criteria take into account the parsimony principle and in a more marked way in the SBIC criterion.

- *Behaviour of the C_2 criterion*

From (17) and (18) we know the expressions for SSD_1 and SSD_2, and applying Lemma 1, we obtain:

$$\frac{SSD_1}{N} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - F(a_i'\hat{\gamma})\right)^2 \xrightarrow{p} \frac{\sum_{i=1}^{N} E_0\left(y_i - F(a_i'\gamma^*)\right)^2}{N} \tag{30}$$

$$\frac{SSD_2}{N} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - F(b_i'\hat{\delta})\right)^2 \xrightarrow{p} \frac{\sum_{i=1}^{N} E_0\left(y_i - F(b_i'\delta^*)\right)^2}{N} \tag{31}$$

Taking into account expressions (20) and (21), we can see that the limits of (30) and (31) are equal, that is to say, $plim\left(\frac{SSD_1}{N}\right) = plim\left(\frac{SSD_2}{N}\right)$; additionally, the terms $\left(1+\frac{2k_j}{N}\right)$ (j = 1,2) converge to 1 when $N \to \infty$, so we can conclude:

$$plim\ [C_2\,(\mathrm{M1})] = plim\ C_2\,[(\mathrm{M2})] \tag{32}$$

meaning that the C_2 criterion behaves well in this context. The converge rate is different between M1 and M2 when $k_1 \neq k_2$. The term $\frac{2k_j}{N}$ is $o\left(N\right)$ and converges to zero faster for models with less k_j.

Conclusion of Alternative 1.1: The three criteria we have analysed satisfy definition 1, meaning their behaviour is good.

3.2. <u>Alternative 1.2</u>

Now the competing models are those given in (7):

M1: $p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 w_i) = F(a_i'\gamma)$

M2: $p_i = F(\delta_0 + \delta_1 x_{1i} + \delta_2 z_i) = F(b_i'\delta)$

As we can see, the family of models containing the DGP is nested in model M1 but not in M2, so the estimated parameter vector $\hat{\gamma}$ converges to the true parameter vector, but not $\hat{\delta}$. In other terms:

$$\hat{\gamma} \xrightarrow{p} \gamma^* = \beta^{0+} \tag{33}$$

$$\hat{\delta} \xrightarrow{p} \delta^* \neq \beta^{0+} \tag{34}$$

so it is obvious that M1 is closer to the DGP than M2.

The following definition establishes how a criterion considered adequate should behave if one of the competing models is better than the other.

DEFINITION 2. Given the models M1 and M2, where M1 is closer to the DGP, a selection criteria (R) is considered as adequate if it satisfies:

$$plim\ [\ \mathrm{R(M1)}\ ] < plim\ [\ \mathrm{R(M2)}\ ] \tag{35}$$

That is to say, if it tends to select M1.

From (4), if M1 is closer to the DGP than M2, then:

$$E\left[\log \frac{f\left(a_i'\gamma^*\right)}{f\left(b_i'\delta^*\right)}\right] > 0 \tag{36}$$

Following a derivation similar to (22), this "superiority condition" of M1 versus M2 can be written as:

$$\left[F_{0i}\log\frac{F_{1i}}{F_{2i}} + (1-F_{0i})\log\frac{1-F_{1i}}{1-F_{2i}}\right] > 0 \tag{37}$$

- *Behaviour of the information criteria*

Following definition 2, the Information Criteria are adequate if:

$$plim\ [\mathrm{IC(M1)}\ ] < plim\ [\mathrm{IC(M2)}] \tag{38}$$

In other terms, we must obtain:

$$plim\left[-\frac{\hat{\ell}_1}{N} + \frac{K_N\left(M1\right)}{N}\right] < plim\left[-\frac{\hat{\ell}_2}{N} + \frac{K_N\left(M2\right)}{N}\right] \tag{39}$$

Given that $\frac{K_N\left(M_j\right)}{N} \xrightarrow{p} 0$, expression (39) will be satisfied if:

$$plim\left(\frac{\hat{\ell}_1}{N}\right) > plim\left(\frac{\hat{\ell}_2}{N}\right) \tag{40}$$

Using the convergence results of (27), we express (40) as follows:

$$\frac{1}{N}\sum_{i=1}^{N} E\left[Y_i \log F_{1i} + \left(1-Y_i\right)\log(1-F_{1i}\right] > \frac{1}{N}\sum_{i=1}^{N} E\left[Y_i \log F_{2i} + \left(1-Y_i\right)\log(1-F_{2i})\right]$$

or

$$\frac{1}{N}\sum_{i=1}^{N}\left[F_{0i}\log\frac{F_{1i}}{F_{2i}} + (1-F_{0i})\log\frac{1-F_{1i}}{1-F_{2i}}\right] > 0 \tag{41}$$

When M1 is closer to the DGP than M2, condition (37) is always satisfied. Now we observe that expression (41) is identical to (37), so we can affirm that the IC criteria satisfy (38), which is the condition for an adequate criterion.

- *Behaviour of the* C_2 *criterion*

The convergences in probability (30) and (31) are satisfied in this case but, according to conditions (33) and (34) ($\gamma^* \neq \delta^*$), $plim\left(\frac{SSD_1}{N}\right) \neq plim\left(\frac{SSD_2}{N}\right)$ and then, $plim\left[C_2\left(M_1\right)\right] \neq plim\left[C_2\left(M_2\right)\right]$. In order to analyse which of the limits is the largest, we focus on expressions (30) and (31).

In (30), we use $E_0\left(y_i\right) = F(x_i'\beta^0)$ and $F(a_i'\gamma^*) = F(x_i'\beta^0)$, resulting in:

$$\frac{\sum_{i=1}^{N} E_0\left(y_i - F_{1i}\right)^2}{N} = \frac{\sum_{i=1}^{N} Var(y_i)}{N} = \frac{\sum_{i=1}^{N} F_{0i}\left(1-F_{0i}\right)}{N} \tag{42}$$

However, the limit given in (31) $F(b_i'\delta^*)$ is different from $F(x_i'\beta^0)$, so we have:

$$\frac{\sum_{i=1}^{N} E_0\left(y_i - F_{2i}\right)^2}{N} = \frac{\sum_{i=1}^{N}\left[E_0(y_i^2) + \left(F_{2i}\right)^2 - 2F_{2i}E_0(y_i)\right]}{N}$$

Given that $E_0\left(y_i^{\,2}\right) = Var\left(y_i\right) + [E_0\left(y_i\right)]^2$, we obtain:

$$\frac{\sum_{i=1}^{N} E_0\left(y_i - F_{2i}\right)^2}{N} = \frac{\sum_{i=1}^{N} F_{0i}\left(1-F_{0i}\right)}{N} + \frac{\sum_{i=1}^{N}\left[F_{0i} - F_{2i}\right]^2}{N} \tag{43}$$

From results (42) and (43), we conclude that:

$$\frac{SSD_1}{N} \xrightarrow{p} \frac{\sum_{i=1}^{N} F_{0i}\left(1-F_{0i}\right)}{N} = h_1$$

$$\frac{SSD_2}{N} \xrightarrow{p} \frac{\sum_{i=1}^{N} F_{0i}\left(1-F_{0i}\right)}{N} + \frac{\sum_{i=1}^{N}\left[F_{0i}-F_{2i}\right]^2}{N} = h_1 + h_2$$

with h_1 and h_2 positive terms, so $plim\left(\frac{SSD_1}{N}\right) < plim\left(\frac{SSD_2}{N}\right)$.

Therefore, taking into account that the terms $\left(1+\frac{2k_j}{N}\right)$ (j = 1,2) converge to 1 when $N \to \infty$, we conclude that

$$plim\left[C_2\left(M_1\right)\right] < plim\left[C_2\left(M_2\right)\right] \tag{44}$$

and that the C_2 criterion satisfies (35), which means it is adequate.

Conclusion of Alternative 1.2: All the criteria satisfy definition 2.

3.3. Alternative 1.3

The competing models are given in (8):

M1: $p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 w_i) = F\,(a_i'\gamma\,)$

M2: $p_i = F(\delta_0 + \delta_1 x_{2i} + \delta_2 w_i) = F\,(b_i'\delta\,)$

Given that the family of models corresponding to the DGP is not nested in any of these two models, the convergence results for the estimates parameter of both models are:

$$\hat{\gamma} \xrightarrow{p} \gamma^* \neq \beta^{0+} \tag{45}$$

$$\hat{\delta} \xrightarrow{p} \delta^* \neq \beta^{0+} \tag{46}$$

Unlike the other scenarios previously considered, this result establishes that there is no theoretical information for knowing if the distance of M1 to the DGP is the same as that of M2 (it would mean that the models are equivalent), or one model is closer to the DGP than the other (which would mean that one model is better than the other).

According to Vuong (1989), there are two possible situations:

a) $f\,(Y_i;\gamma^*,a_i) = f\,(Y_i;\delta^*,b_i)$, that is to say, the density functions of Y_i in M1 and M2 evaluated at the corresponding pseudo-true parameter vectors are observationally identical. As Vuong (1989) shows, it implies that M1 and M2 are equivalent specifications.

b) $f(Y_i;\gamma^*,a_i) \neq f(Y_i;\delta^*,b_i)$. In these situations the models can be equivalent or non-equivalent, which means that $E_0\left[\ell_1^*\right]=E_0\left[\ell_2^*\right]$ or $E_0\left[\ell_1^*\right]\neq E_0\left[\ell_2^*\right]$, respectively.

The definition of an adequate selection criterion depends on the specific scenario. So, if the models are equivalent, we must give credence to definition 1 (expression (24)), while we will use definition 2 (expression (35)) if one of the models is closer to the DGP than the other one. Now, we analyse each of these situations.

a) <u>The density function at the pseudo-true parameters are observationally identical</u> ($f(Y_i;\gamma^*,a_i) = f(Y_i;\delta^*,b_i)$**)**.

In general terms the density function of a binomial variable, like Y_i is $f(Y_i)=\left(p_i\right)^{Y_i}\left(1-p_i\right)^{1-Y_i}$**.** Then, for Y_i=1, the density is $p_i\,(=F\left(\bullet\right)$**)**, and if Y_i= 0, it is (1- p_i). It means that the densities of Y_i in M1 and M2 will be equal if $F(a_i'\gamma^*)=F(b_i'\delta^*)$**.** The convergence results (45) and (46) show that $\gamma^*\neq\delta^*$ and both are different from β^0, so $a_i'\gamma^*=b_i'\delta^*$ must be true to arrive at the equality $F(a_i'\gamma^*)=F(b_i'\delta^*)$.

We now analyse whether or not the criteria satisfy definition 1.

- *Behaviour of the IC criteria*

We know that M1 and M2 are equivalent, that is to say, $E_0\left[\ell_1^*\right]=E_0\left[\ell_2^*\right]$ which means that expression (23) is always satisfied. This situation is similar to that of alternative 1.1, with the only difference being that now $\gamma^*\neq\delta^*$. In spite of this inequality, to arrive at $plim\left(\mathrm{IC}\left(\mathrm{M1}\right)\right)=plim\left(\mathrm{IC}\left(\mathrm{M2}\right)\right)$ expression (28) must hold, which is true because (28) and (23) are identical.

- *Behaviour of the C_2 criterion.*

Application of Lemma 1 establishes that convergences given in (30) and (31) hold, but under (45) and (46) and using a development similar to that used to obtain (43), we have:

$$\frac{\sum_{i=1}^{N}E\left(y_i-F(a_i'\gamma^*)\right)^2}{N}=\frac{\sum_{i=1}^{N}F_{0i}\left(1-F_{0i}\right)}{N}+\frac{\sum_{i=1}^{N}\left[F_{0i}-F_{1i}\right]^2}{N}$$

$$\frac{\sum_{i=1}^{N} E\left(y_i - F(b_i'\delta^*)\right)^2}{N} = \frac{\sum_{i=1}^{N} F_{0i}\left(1 - F_{0i}\right)}{N} + \frac{\sum_{i=1}^{N}\left[F_{0i} - F_{2i}\right]^2}{N}$$

The convergence limits in this case are:

$$\frac{SSD_1}{N} \xrightarrow{p} \frac{\sum_{i=1}^{N} F_{0i}\left(1 - F_{0i}\right)}{N} + \frac{\sum_{i=1}^{N}\left[F_{0i} - F_{1i}\right]^2}{N} = h_1 + h_3 \qquad (47)$$

$$\frac{SSD_2}{N} \xrightarrow{p} \frac{\sum_{i=1}^{N} F_{0i}\left(1 - F_{0i}\right)}{N} + \frac{\sum_{i=1}^{N}\left[F_{0i} - F_{2i}\right]^2}{N} = h_1 + h_2 \qquad (48)$$

h_i (i = 1, 2, 3) being positive constants.

Now, we need to know which of the two previous convergence limits (47) and (48) has the largest value, in other words, the relationship between h_3 and h_2.

Given that the starting point is the equivalence of M1 and M2, we should have $h_3 = h_2$ in order to obtain adequate behaviour of the criterion. As we commented previously, the equality of the density functions at the pseudo-true parameters makes $F_{1i} = F_{2i}$, so we can affirm that $h_3 = h_2$.

Then we can conclude that $plim\left[C_2\left(M_1\right)\right] = plim\left[C_2\left(M_2\right)\right]$ meaning this criterion performs adequately.

<u>b) The density function at the pseudo-true parameters are observationally different</u> ($f(Y_i;\gamma^*, a_i) \neq f(Y_i;\delta^*, b_i)$).

Similarly to a), we conclude that $F(a_i'\gamma^*) \neq F(b_i'\delta^*)$. Under this condition, the two following situations can be found.

b.1) Equivalent models.

In this case, $E\left[\ell_1^*\right] = E\left[\ell_2^*\right]$, which means that equation (23) is always satisfied.

- *Behaviour of the IC criteria*

We are again in a situation similar to alternative 1.1. As we have already shown, the condition for concluding that the IC criteria are adequate (expression (28)) is the same as expression (23), which defines equivalent models. Therefore, we see that the fact that $\gamma^* \neq \delta^*$, and even that $F(a_i'\gamma^*) \neq F(b_i'\delta^*)$ doesn't affect the results.

- *Behaviour of the C_2 criterion.*

We know $F\left(a_i'\gamma^*\right) \neq F\left(b_i'\delta^*\right)$ as well as the probability limits $plim\left(\frac{SSD_1}{N}\right)$ and $plim\left(\frac{SSD_2}{N}\right)$, which were given in expressions (47) and (48). Analogous to alternative 1.3.a, to obtain adequate criterion we must conclude that $h_3 = h_2$, or in other terms

$$\frac{\sum_{i=1}^{N}\left[F_{0i}-F_{1i}\right]^2}{N} = \frac{\sum_{i=1}^{N}\left[F_{0i}-F_{2i}\right]^2}{N} \tag{49}$$

The only possibility for achieving equality (49) is that, on average, $\left[F_{0i}-F_{1i}\right] = -\left[F_{0i}-F_{2i}\right]$, which is equivalent to $2F_{0i} = F_{1i} + F_{2i}$. If this happens, the C_2 criterion will behave correctly. The Monte Carlo experiment will allow us a more specific analysis of the behaviour of the criterion.

b.2) Non-equivalent models.

- *Behaviour of the IC criteria*

Assuming that M1 is closer to the DGP than M2, results (36) and (37) are satisfied. If the IC criteria are adequate, definition 2 establishes that *plim* [IC(M1)] < *plim* [IC(M2)], which is given in (38). The development is identical to that of alternative 1.2, with the difference that now neither γ^* nor δ^* converge to β^0. In spite of this difference, we can conclude that the IC criteria behave correctly.

- *Behaviour of the C_2 criterion.*

We know that $F(a_i'\gamma^*) \neq F(b_i'\delta^*)$, and taking the probability limits obtained in (47) and (48), we conclude that if C_2 behaves well, *plim*[C_2(M1)] < *plim*[C_2(M2)], which means that $h_3 < h_2$. Then, we must prove:

$$\frac{\sum_{i=1}^{N}\left[F_{0i}-F_{1i}\right]^2}{N} < \frac{\sum_{i=1}^{N}\left[F_{0i}-F_{2i}\right]^2}{N} \tag{50}$$

Or equivalently:

$$\frac{\sum_{i=1}^{N}\left[F_{0i}-F_{1i}\right]^2}{N} - \frac{\sum_{i=1}^{N}\left[F_{0i}-F_{2i}\right]^2}{N} < 0 \tag{51}$$

We analyse the behaviour of a "typical individual", and afterwards we will represent an "average behaviour".

Given that M1 is closer to the DGP than M2, we use the equation:

KLD (M0,M1) < KLD(M0,M2)

which, from (4) we can express as:

$$E_0\,[\,\ell_2\, - \,\ell_1\,] = E_0\,[\,\ell_2\,] - E_0\,[\,\ell_1\,] < 0$$

that leads to expression (37). If we eliminate the logarithm of (37), the superiority condition of M1 to M2 results in:

$$\left(\frac{F_{1i}}{F_{2i}}\right)^{F_{0i}} \cdot \left(\frac{1-F_{1i}}{1-F_{2i}}\right)^{1-F_{0i}} > 1 \Rightarrow \left(\frac{F_{1i}}{F_{2i}}\right)^{F_{0i}} > \left(\frac{1-F_{2i}}{1-F_{1i}}\right)^{1-F_{0i}} \tag{52}$$

We want to know the combinations of F_{0i}, F_{1i} and F_{2i} which allows (52) to be satisfied. First, we will study the case where F_{0i} has the value 0, 0.5 and 1, and afterwards we will analyze what happens in the intervals (0, 0.5) and (0.5, 1).

⇒ F_{0i} = 0.

Expression (52) becomes:

$$1 > \frac{1-F_{2i}}{1-F_{1i}} \Rightarrow (1 - F_{1i}) > (1 - F_{2i}) \Rightarrow F_{1i} < F_{2i}$$

concluding that if F_{0i} = 0 and M1 is better than M2, the following inequality is satisfied:

$$F_{0i} < F_{1i} < F_{2i} \tag{53}$$

⇒ F_{0i} =1.

In this case (52) can be written as:

$$\frac{F_{1i}}{F_{2i}} > 1 \Rightarrow F_{1i} > F_{2i}$$

and we can conclude that, for F_0 = 1 and M1 better than M2, the following inequality holds:

$$F_{0i} > F_{1i} > F_{2i} \tag{54}$$

⇒ F_{0i} = 0.5.

In condition (52), both sides of the inequality have the same exponent, enabling the expression:

$$\frac{F_{1i}}{F_{2i}} > \frac{1-F_{2i}}{1-F_{1i}} \Rightarrow F_{1i}\ (1 - F_{1i}) > F_{2i}\ (1 - F_{2i}) \Rightarrow F_{1i}^2 - F_{2i}^2 < F_{1i} - F_{2i}$$

$$\Rightarrow (F_{1i} - F_{2i})\ (F_{1i} + F_{2i}) < F_{1i} - F_{2i} \tag{55}$$

Inequality (55) holds in two situations:

a) If $F_{1i} > F_{2i}$ and $(F_{1i} + F_{2i}) < 1$ (56)

b) If $F_{1i} < F_{2i}$ and $(F_{1i} + F_{2i}) > 1$ (57)

so if F_{0i} = 0.5 and M1 is better than M2, one of the two inequalities, (56) or (57), will be satisfied.

Now, for the three situations considered, we can verify if the C_2 criterion performs well, that is to say, if expression (51) is hold. First, (53) and (54) lead to the conclusion that $h_3 < h_2$, so C_2 satisfies definition 2. Secondly, if F_{0i} = 0,5 and (56) or (57) holds, the development of (51) allows us to obtain:

$$\frac{\sum_{i=1}^{N}\left[(F_{1i} - F_{2i})\,[F_{1i} + F_{2i} - 2F_{0i}]\right]}{N} < 0 \tag{58}$$

or equivalently:

$$\frac{\sum_{i=1}^{N}\left[(F_{1i} - F_{2i})\,[F_{1i} + F_{2i} - 1]\right]}{N} < 0 \tag{59}$$

If (56) or (57) holds, then (59) is verified and definition 2 is again satisfied. In summary, we can say that when F_{0i} adopts values 0, 0.5 or 1, the C_2 criterion establishes that model M1 is closer to DGP than model M2, which is the correct result.

The study of the superiority condition of M1 when F_{0i} $\in$ (0, 0.5) or (0.5, 1) is developed in Appendix 2. The obtained results, that is to say, the combinations (F_{0i}, F_{1i}, F_{2i}) which satisfies the superiority condition are:

$$\rightarrow F_{0i} \in (0,\, 0.5) \begin{cases} F_{1i} < F_{2i} & y \quad F_{1i} + F_{2i} > 1 \\ F_{1i} > F_{2i} & y \quad F_{1i} + F_{2i} < 1 \quad y \quad F_{0i} \cong 0.5 \\ F_{1i} < F_{2i} & y \quad F_{1i} + F_{2i} < 1 \quad y \quad F_{0i} \cong 0 \end{cases} \tag{60}$$

$$\rightarrow F_{0i} \in (0.5,\, 1) \begin{cases} F_{1i} > F_{2i} & y \quad F_{1i} + F_{2i} < 1 \\ F_{1i} > F_{2i} & y \quad F_{1i} + F_{2i} > 1 \quad y \quad F_{0i} \cong 1 \\ F_{1i} < F_{2i} & y \quad F_{1i} + F_{2i} > 1 \quad y \quad F_{0i} \cong 0.5 \end{cases} \tag{61}$$

For every combination of (60) and (61) the following is obtained:

$$plim\,[C_2\ \text{(M1)} - C_2\ \text{(M2)}] = \frac{\sum_{i=1}^{N}\left[(F_{1i} - F_{2i})\,[F_{1i} + F_{2i} - 2F_{0i}]\right]}{N} < 0$$

so, the C_2 criterion behaves adequately, according to definition 2.

Conclusion of Alternative 1.3: The study of all the situations included in this alternative leads us to conclude that the three selection criteria perform well. The only exception is the behaviour of C_2 for equivalent models with non-identical densities at the pseudo-true parameters. In this case, such criterion behaves adequately in only specific situations.

3.4. Case 2

Turning to (9), the competing models are:

M1: $p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}) = F\,(a_i'\gamma)$

M2: $p_i = F(\delta_0 + \delta_1 x_{1i} + \delta_2 z_i) = F\,(b_i'\delta)$

so M1 belongs to the model family that includes the DGP; in other words, M1 is correctly specified.

In this situation, the convergence results for the estimated parameters are:

$$\hat{\gamma} \xrightarrow{p} \gamma^* = \beta^{0+} \qquad (62)$$

$$\hat{\delta} \xrightarrow{p} \delta^* \neq \beta^{0+} \qquad (63)$$

Under this framework, a given criterion will be adequate if definition 2 is satisfied, or equivalently, if the criterion tends to select M1.

This case is similar to alternative 1.2 when M1 converges to the DGP while M2 is further from it. Then, the results we obtain are the same as that of this alternative.

Conclusion of Case 2: The three criteria perform well.

The theoretical results we have obtained are summarised in Table 1.

Table 1: Summary of the behaviour of the criteria.

Relationship between DGP and M*j* (*j* = 1, 2)	**Relationship between M1 and M2**	**Behaviour**	
		IC	C_2
DGP nested in M*j* (*j* = 1, 2) (*Case1*, *Alternative 1.1*)	Equivalent	*plim* [IC(M1) –IC(M2)] = 0 Adequate	*plim*[C_2 (M1)– C_2 (M2)] = 0 Adequate
DGP only nested in M1 (*Case1*, *Alternative 1.2)*	M1 better than M2	*plim* [IC(M1) –IC(M2)] < 0 Adequate	*plim*[C_2 (M1)– C_2 (M2)] < 0 Adequate
DGP not nested in M*j* (*j* = 1, 2) (*Case1*, *Alternative 1.3*)	Identical densities at the pseudo-trues $\Rightarrow$ Equivalent models. *(Situation a)*	*plim* [IC(M1) –IC(M2)] = 0 Adequate	*plim*[C_2 (M1)– C_2 (M2)] = 0 Adequate
	Non identical densities at the pseudo-trues, but equivalent models. *(Situation b.1)*	*plim* [IC(M1) –IC(M2)] = 0 Adequate	*plim* [C_2 (M1) – C_2 (M2)] = 0 Adequate in some cases
	Non identical densities at the pseudo-trues and M1 better than M2. *(Situation b.2)*	*plim* [IC(M1) –IC(M2)] < 0 Adequate	*plim*[C_2 (M1)– C_2 (M2)] < 0 Adequate
DGP $\in$ M1 *(Case 2)*	M1 correctly specified	*plim* [IC(M1) –IC(M2)] < 0 Adequate	*plim*[C_2 (M1)– C_2 (M2)] < 0 Adequate

4. MONTE CARLO EXPERIMENTS

The objective of the Monte Carlo experiments is twofold: confirm the theoretical results and study the behaviour of all criteria in finite samples.

Within our framework, the models we specify attempt to explain binary variable y_i, and for every estimated model we obtain the numerical value of the selection criteria which were theoretically analyzed in the previous sections. Additionally, we must specify the DGP which generated the values of the response variable y_i. The generation of this variable is based on the latent linear model which underlies any binary model:

$$y_i^* = x_i'\beta + u_i \qquad (64)$$

where y_i^* is a latent variable (not observable) and depends on y_i through:

$$y_i = \begin{cases} 1 & if \quad y_i^* > 0 \\ 0 & if \quad y_i^* \leq 0 \end{cases} \qquad (65)$$

Under assumption (A.1) of Section 3, and following the procedure of Gourieroux *et al.* (1987), we obtained the values of y_i. We considered different sets of parameter values and different kinds of explanatory variables (continuous and dummy). We chose the standard normal distribution function for the error term implying exclusive focus on probit models. We established two sample sizes N = 200 and 2000 and we carried out 500 replications (R) for each experiment. In each replication, we changed the error term to generate different series of the latent variable underlying the response variable. Additionally, we fixed the intercept at a value of -2, which allowed minimize the troubles associated to a non-adequate number of ones in the sample of y_i; too few or too many ones can lead to problems when estimating and interpreting results.

The experiments attempt to verify the theoretical results summarised in Table 1. We estimated M1 and M2 500 times (there are 500 samples of y_i) for every experiment. The corresponding tables present the number of times that each criterion selects M1.

MONTECARLO EXERCICE FOR CASE 1, ALTERNATIVE 1.1

The first set of experiments is focused on alternative 1.1, where the DGP is nested in both competing models, M1 and M2. We distinguished two scenarios in order to take into account that the two modes can include different numbers of explanatory variables. The framework is:

Scenario A: $$\begin{aligned} M1 &\equiv p_i = F\left(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 w_i\right) \\ M2 &\equiv p_i = F\left(\delta_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \delta_3 z_i\right) \end{aligned}$$

Scenario B: $$\begin{aligned} M1 &\equiv p_i = F\left(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 w_i + \gamma_4 s_i\right) \\ M2 &\equiv p_i = F\left(\delta_0 + \delta_1 x_{1i} + \delta_2 x_{2i} + \delta_3 z_i\right) \end{aligned}$$

PGD: $p_i = F\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right)$

For variables x_1 and x_2 we considered the following combinations:

i) x_1, x_2 ~ U(0,1), ii) x_1 ~ U(0,1), $x_2 \sim \chi_1^2$, iii) x_1, x_2 ~ χ_1^2 , iv) x_1 ~ χ_1^2 x_2~dummy. Variables *w*, *z* and *s* are generated as U(0,1), N(3,1) and χ^2, respectively.

Tables 2 and 3 present the results for N = 200 and 2000, respectively.

As we can summarize in Table 1, models M1 and M2 are equivalent, so a given criterion behaves well if it is indifferent to selecting one model or the other one. This means that the number presented in each cell of Tables 2 and 3 should be around 250 (50% of 500 times). This is the behaviour we observe for scenario A for each criterion, and for every sample size considered. However, if we consider Scenario B, where M1 has a higher number of irrelevant regressors than M2, we can see that the criteria tend to select the model with fewer parameters, that is to say, it selects the more parsimonious model (M2), and this tendency grows with the sample size. The most parsimonious criterion is SBIC, followed by AIC, and finally C_2. This characteristic was explained in the previous section, so the Monte Carlo exercise corroborates this theoretical aspect. .

Neither the kind of variables (uniform, chi-squared, normal or dummy), nor the set of values of the DGP parameter vector seem to affect the behaviour of the criteria.

Table 2: *Behaviour of the selection criteria in Alternative 1.1.* N = 200

	DGP			Specified model					
	β'	x_1	x_2	w	**z**	s	**AIC**	**SBIC**	C_2
Scenario A	(-2,3,1)	U	U	U	N(3,1)		240	240	247
	(-2,1,3)	U	U	U	N(3,1)		259	259	252
	(-2,1,1)	U	U	U	N(3,1)		233	233	232
	(-2,3,1)	χ^2	χ^2	U	N(3,1)		252	252	240
	(-2,1,3)	χ^2	χ^2	U	N(3,1)		251	251	257
	(-2,1,1)	χ^2	χ^2	U	N(3,1)		249	249	270
	(-2,3,1)	U	χ^2	U	N(3,1)		252	252	256
	(-2,1,3)	U	χ^2	U	N(3,1)		256	256	245
	(-2,1,1)	U	χ^2	U	N(3,1)		247	247	243
	(-2,3,1)	U	*dm*	U	N(3,1)		240	240	263
	(-2,1,3)	U	*dm*	U	N(3,1)		234	234	235
	(-2,1,1)	U	*dm*	U	N(3,1)		235	235	231
	(-2,3,1)	χ^2	*dm*	U	N(3,1)		242	242	248
	(-2,1,3)	χ^2	*dm*	U	N(3,1)		260	260	257
	(-2,1,1)	χ^2	*dm*	U	N(3,1)		249	249	249
Scenario B	(-2,3,1)	U	U	U	N(3,1)	χ^2	128	21	154
	(-2,1,3)	U	U	U	N(3,1)	χ^2	134	29	165
	(-2,1,1)	U	U	U	N(3,1)	χ^2	139	24	165
	(-2,3,1)	χ^2	χ^2	U	N(3,1)	χ^2	128	28	212
	(-2,1,3)	χ^2	χ^2	U	N(3,1)	χ^2	136	35	209
	(-2,1,1)	χ^2	χ^2	U	N(3,1)	χ^2	154	28	214
	(-2,3,1)	U	χ^2	U	N(3,1)	χ^2	137	29	195
	(-2,1,3)	U	χ^2	U	N(3,1)	χ^2	136	26	208
	(-2,1,1)	U	χ^2	U	N(3,1)	χ^2	136	28	190
	(-2,3,1)	U	*dm*	U	N(3,1)	χ^2	119	34	158
	(-2,1,3)	U	*dm*	U	N(3,1)	χ^2	137	27	164
	(-2,1,1)	U	*dm*	U	N(3,1)	χ^2	121	33	164
	(-2,3,1)	χ^2	*dm*	U	N(3,1)	χ^2	140	34	200
	(-2,1,3)	χ^2	*dm*	U	N(3,1)	χ^2	144	29	219
	(-2,1,1)	χ^2	*dm*	U	N(3,1)	χ^2	125	28	180

Table 3: *Behaviour of the selection criteria in Alternative 1.1*. N = 2000

	DGP			Specified model					
	β'	x_1	x_2	w	**z**	s	**AIC**	**SBIC**	C_2
Scenario A	(-2,3,1)	U	U	U	N(3,1)		273	273	275
	(-2,1,3)	U	U	U	N(3,1)		236	236	240
	(-2,1,1)	U	U	U	N(3,1)		234	234	244
	(-2,3,1)	χ^2	χ^2	U	N(3,1)		240	240	252
	(-2,1,3)	χ^2	χ^2	U	N(3,1)		273	273	261
	(-2,1,1)	χ^2	χ^2	U	N(3,1)		236	236	238
	(-2,3,1)	U	χ^2	U	N(3,1)		247	247	245
	(-2,1,3)	U	χ^2	U	N(3,1)		231	231	245
	(-2,1,1)	U	χ^2	U	N(3,1)		243	243	260
	(-2,3,1)	U	*dm*	U	N(3,1)		259	259	269
	(-2,1,3)	U	*dm*	U	N(3,1)		231	231	245
	(-2,1,1)	U	*dm*	U	N(3,1)		242	242	261
	(-2,3,1)	χ^2	*dm*	U	N(3,1)		255	255	251
	(-2,1,3)	χ^2	*dm*	U	N(3,1)		248	248	257
	(-2,1,1)	χ^2	*dm*	U	N(3,1)		251	251	234
Scenario B	(-2,3,1)	U	U	U	N(3,1)	χ^2	144	10	179
	(-2,1,3)	U	U	U	N(3,1)	χ^2	133	10	151
	(-2,1,1)	U	U	U	N(3,1)	χ^2	126	2	144
	(-2,3,1)	χ^2	χ^2	U	N(3,1)	χ^2	128	12	215
	(-2,1,3)	χ^2	χ^2	U	N(3,1)	χ^2	140	10	230
	(-2,1,1)	χ^2	χ^2	U	N(3,1)	χ^2	130	11	185
	(-2,3,1)	U	χ^2	U	N(3,1)	χ^2	128	9	165
	(-2,1,3)	U	χ^2	U	N(3,1)	χ^2	106	2	192
	(-2,1,1)	U	χ^2	U	N(3,1)	χ^2	120	5	161
	(-2,3,1)	U	*dm*	U	N(3,1)	χ^2	144	12	177
	(-2,1,3)	U	*dm*	U	N(3,1)	χ^2	145	7	153
	(-2,1,1)	U	*dm*	U	N(3,1)	χ^2	141	7	162
	(-2,3,1)	χ^2	*dm*	U	N(3,1)	χ^2	147	10	200
	(-2,1,3)	χ^2	*dm*	U	N(3,1)	χ^2	136	8	185
	(-2,1,1)	χ^2	*dm*	U	N(3,1)	χ^2	143	12	185

MONTECARLO EXERCISE FOR CASE 1, ALTERNATIVE 1.2

$$M1 \equiv p_i = F\left(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 w_i\right)$$

$$M2 \equiv p_i = F\left(\delta_0 + \delta_1 x_{1i} + \delta_2 z_i\right)$$

PGD: $p_i = F\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right)$

Tables 4 and 5 summarise the results of this situation. A criterion behaves well if it almost always selects M1 (approximately 500 times). When N = 2000 the theoretical results are corroborated, so all the criteria tend to select M1, whatever the kind of explanatory variables. However, for N = 200 we can see that the results are not so evident, although they tend to adequate behaviour. Specifically, we can observe that differences are found when the variables x_1 and x_2 are both uniform and the weight of x_2 is not greater than that of x_1 and this difference is more evident for the SBIC criterion. When the DGP contains a dummy variable, the results are in accordance with theoretical derivations.

Table 4: *Behaviour of the selection criteria for Alternative 1.2*. N = 200

DGP			Specified Model				
β'	x_1	x_2	w	z	AIC	SBIC	C_2
(-2,3,1)	U	U	U	N(3,1)	462	373	458
(-2,1,3)	U	U	U	N(3,1)	500	500	500
(-2,1,1)	U	U	U	N(3,1)	436	329	433
(-2,3,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	*dm*	U	N(3,1)	500	497	500
(-2,1,3)	U	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	U	*dm*	U	N(3,1)	498	495	497
(-2,3,1)	χ^2	*dm*	U	N(3,1)	495	462	493
(-2,1,3)	χ^2	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	*dm*	U	N(3,1)	497	489	498

Table 5 *Behaviour of the selection criteria for Alternative1.2*.N = 2000

PGD			Specified Model				
β'	x_1	x_2	w	**z**	**AIC**	**SBIC**	C_2
(-2,3,1)	U	U	U	N(3,1)	500	500	500
(-2,1,3)	U	U	U	N(3,1)	500	500	500
(-2,1,1)	U	U	U	N(3,1)	500	500	500
(-2,3,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	*dm*	U	N(3,1)	500	500	500
(-2,1,3)	U	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	U	*dm*	U	N(3,1)	500	500	500
(-2,3,1)	χ^2	*dm*	U	N(3,1)	500	500	500
(-2,1,3)	χ^2	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	*dm*	U	N(3,1)	500	500	500

MONTECARLO EXERCICE FOR CASE 1, ALTERNATIVE 1.3

This is the most complicated framework because the DGP is not nested in any of the two competing models.

$$M1 \equiv p_i = F\left(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 w_i\right)$$
$$M2 \equiv p_i = F\left(\delta_0 + \delta_1 x_{2i} + \delta_2 w_i\right)$$

$$\text{PGD: } p_i = F\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right)$$

As we established in the previous section, in this context one of the following two situations can occur:

a) $f(Y_i;\gamma^*,a_i) = f(Y_i;\delta^*,b_i)$, which implies that $E\left(\ell_1^*\right) = E\left(\ell_2^*\right)$, that is to say, both models are equivalent.

b) $f(Y_i;\gamma^*,a_i) \neq f(Y_i;\delta^*,b_i)$, where it is possible that $E\left(\ell_1^*\right) = E\left(\ell_2^*\right)$ or $E\left(\ell_1^*\right) \neq E\left(\ell_2^*\right)$, that is to say, equivalent models or non-equivalent models.

The implementation of this set of experiments should be based on the relationship between the true parameter vector (β^0) and each of the pseudo-true parameter vectors (γ^* and δ^*). Those relationships are derived in Appendix 3. We should be able to obtain the value of β^0

from them which satisfies the equality of density functions. In other terms, if we have $\gamma^* = m_1(\beta^0)$ and $\delta^* = m_2(\beta^0)$, we are interested in obtaining β^0 that makes $f(Y_i; m_1(\beta^0), a_i) = f(Y_i; m_2(\beta^0), b_i)$. The same idea could be used in situation b) in order to equal the log-likelihoods $E(\ell_1^*)$ and $E(\ell_2^*)$, that is to say, $E\left[\ell_1\left(m_1\left(\beta^0\right)\right)\right] = E\left[\ell_2\left(m_2\left(\beta^0\right)\right)\right]$. Nevertheless, the non-linear equation system we should solve has insurmountable problems. We needed to find a strategy which allows us to approximate the previous equalities of densities and likelihoods. To this end, we generated several DGP modifying the value of the parameter vector β^0 and the kind of explanatory variables. Again, the intercept was fixed at a value of -2, while β_1 and β_2 took values in the range of (-2, 2) counting by 0.5s, and additionally they took values 3, 4, 5 and 7. As a result of this strategy, we generated 132 different DGPs (once we eliminated the repeated combinations). Each of the outlined DGPs led to a specific relationship between models M1 and M2: equivalent models (with identical or non-identical densities) or non-equivalent models.

First, in order to distinguish the cases where the models are equivalent from those which are not, for each of the N observations corresponding to a given generation, we made the following calculation:

$$del_i = E(\ell_{1i}^*) - E(\ell_{2i}^*) \qquad i = 1, \cdots, N$$

To obtain the values of del_i we needed the relationship between true and pseudo-true parameters. This relationship depends only on the explanatory variables, so the expression was the same for the 500 replications.

Models M1 and M2 will be equivalents if $E(\ell_{1i}^*) = E(\ell_{2i}^*)$ $\forall i$, that is to say, if $del_i = 0$ $\forall i$ =0**.** Nevertheless, it is very difficult to obtain, so we tried to find the proximity to the equivalence. In other terms, $E(\ell_{1i}^*) - E(\ell_{2i}^*) \cong 0$, for the majority of the N observations. Then, this information regarding the N observations was summarised as follows:

$$absdel = \frac{1}{N}\sum_{I=1}^{n}\left|E(\ell_{1i}^*) - E(\ell_{2i}^*)\right|$$

The absolute value prevents the observations from balancing each other. The closest to zero value was this measurement, the closest to being equivalent are the competing models. But, could we use only the *absdel* measure to affirm that the models are equivalents? The answer is no, because we can obtain *absdel* $\cong 0$ but with the majority of the observations satisfying $E(\ell_1^*) > E(\ell_2^*)$, which means that M1 is closer to the DGP.

We denoted AM1 the number of observations where $E(\ell_{1i}^*) > E(\ell_{2i}^*)$, that is to say, where model M1 is closer to DGP. Under the assumption that M1 and M2 are equivalent models, the value of AM1 should tend to N/2. Could we use only AM1 to classify the models, without taking into account the values of *absdel*? The answer is no again, because we can obtain AM1 $\cong$ N/2 but with a large *absdel* value due to large differences in $|E(\ell_{1i}^*) - E(\ell_{2i}^*)|$. This argument allowed us to conclude that there are two requirements for establishing whether the models tend to be equivalent or not: a near zero *absdel* value, and an AMI value of close to 50% of the observations (N/2).

Once we established the set of equivalent specifications, we analysed whether some of them have observationally identical densities at the pseudo-true parameter vector. To do this, we used the following measure:

$$difden = \frac{1}{N}\sum_{i=1}^{N}\left|difden_i\right|$$

with $difden_i = f\left(y_i;\gamma*,x_1,x_2\right) - f\left(y_i,\delta^*,x_1,x_2\right)$. Values of *difden* close to zero indicate that the competing models tend to be equivalent with identical densities at the pseudo-trues.

We then focused on the non-equivalent models, in order find out which of them is the closest to the DGP in each case. Given that AM1 denotes the number of observations where M1 is better than M2, we could use this measurement to conclude which is the best model.

Specifically, in the experiments where N/2 < AM1 < N , M1 could be considered better than M2, while M2 is better if 0 < AM1 < N/2.

We began by carrying out the strategy described above for the sample size of 2000, because the method is based on asymptotic results. As such, we have ordered the 132 experiments from the lowest to the highest value of *absde*l. Table 6 presents the experiments with the lowest value of this measurement. Additionally AM1 and *absdifden* were included. Table 7 shows the experiments where the value of AM1 is close to 1000 (N/2).

Table 6: *Experiments with the lowest value of absdel*. N = 2000.

	DGP					
Experiment	β'	x_1	x_2	**absdel**	AM1	absdifden
3	(-2,1,-1)	U	U	**0,00716535**	1001	0,0227
27	(-2,-1,1)	U	U	**0,00716886**	1003	0,0224
28	(-2,-0.5,1)	U	U	**0,00731291**	505	0,0257
4	(-2,1,-0.5)	U	U	**0,00737039**	1498	0,0261
26	(-2,-1.5,1)	U	U	**0,00989554**	1317	0,0217
2	(-2,1,-1.5)	U	U	**0,01003989**	673	0,0222
25	(-2,-2,1)	U	U	**0,01272868**	1469	0,0218
5	(-2,1,0.5)	U	U	**0,01279404**	1514	0,0513
29	(-2,0.5,1)	U	U	**0,0127966**	510	0,0510
1	(-2,1,-2)	U	U	**0,01315248**	511	0,0226
48	(-2,1,-0.5)	U	χ^2	**0,01515962**	1035	0,0325
6	(-2,1,1)	U	U	**0,01854382**	1000	0,0796
47	(-2,1,-1)	U	χ^2	**0,02005628**	786	0,0341

Table 7: *Experiments with* AM1 *around* 1000. N = 2000

	DGP					
Experiment	β'	x_1	x_2	absdel	**AM1**	absdifden
57	(-2,3,-2)	U	χ^2	0,17489579	**1011**	0,1956
27	(-2,-1,1)	U	U	0,00716886	**1003**	0,0224
3	(-2,1,-1)	U	U	0,00716535	**1001**	0,0227
6	(-2,1,1)	U	U	0,01854382	**1000**	0,0796
63	(-2,3,1.5)	U	χ^2	0,25329468	**995**	0,2597
21	(-2,3,3)	U	U	0,13463303	**991**	0,1867
109	(-2,3,3)	χ^2	χ^2	0,44042798	**980**	0,3307

According to the strategy previously defined, and considering both tables together, experiments 27, 3 and 6 seem to include equivalent models because the values of *absdel* and AM1 are close to zero and 1000, respectively. The rest of the experiments included in tables 6 and 7 only satisfy one of the requirements, so we can’t say that the models are equivalent, but one better than the other.

Once we have found the experiments with equivalent models, we wanted to find out if some of them have identical densities at the pseudo-trues. To do that we sequenced the experiments according to the value of *absdifden*. The smallest values of the measure should correspond to some of the experiments that we previously classified as equivalent; a value of *absdifden* close to zero means directly equivalent models. Then, if the smallest value of *absdifden* does not correspond to cases 27, 3 or 6, the strategy we used is not adequate. Table 8 gives the 13 experiments with the lowest *absdifden* value, and we can see that the three first cases are “non-equivalent models”. Apparently it is a problem of our strategy. Nevertheless, if we calculate the number of ones in the sample of y_i, we find that experiments 1, 2, 25 and 26 have, on average, a very low number of ones, even less than 2.5%. We know that it could be a problem for the estimation, so we had to be careful about the conclusions in such cases. If we omit these four cases, experiments 27 and 3 are in the first places, so we can conclude that the competing models are equivalent with identical densities at the pseudo-true parameter vector. However, experiment 6 has a higher value of *absdifden*, so it is a case of equivalent models with non-identical densities.

Table 8: *Experiments with the lowest absdifden values*. N = 2000.

	PGD					
Experiment	β'	x_1	x_2	absdel	AM1	**Absdifden**
26	(-2,-1.5,1)	U	U	0,00989554	1317	**0,0217**
25	(-2,-2,1)	U	U	0,01272868	1469	**0,0218**
2	(-2,1,-1.5)	U	U	0,01003989	673	**0,0222**
27	(-2,-1,1)	U	U	0,00716886	1003	**0,0224**
1	(-2,1,-2)	U	U	0,01315248	511	**0,0226**
3	(-2,1,-1)	U	U	0,00716535	1001	**0,0227**
28	(-2,-0.5,1)	U	U	0,00731291	505	**0,0257**
4	(-2,1,-0.5)	U	U	0,00737039	1498	**0,0261**
48	(-2,1,-0.5)	U	χ^2	0,01515962	1035	**0,0325**
47	(-2,1,-1)	U	χ^2	0,02005628	786	**0,0341**
46	(-2,1,-1.5)	U	χ^2	0,02293526	668	**0,0349**
45	(-2,1,-2)	U	χ^2	0,02491364	589	**0,0354**
29	(-2,0.5,1)	U	U	0,0127966	510	**0,0510**

Then we needed to classify the rest of the experiments, which correspond to non-equivalent models and with this aim it was necessary to know which was the best model (M1 or M2) for each experiment. To find out it we used the AM1 measure. Table 9 shows all the experiments sequenced accordingly, from the highest to the lowest value of AM1.

We can see that the first part of the table corresponds to experiments where M1 is better, followed by those where both models are equivalent, and finally those where M2 is better. We can observe the changes from one situation to the other one. The column titled *x* indicates the kind of DGP explanatory variables, which are denoted by A, B and C corresponding to the following combinations:

$$A \rightarrow x_1 \sim U(0,1) \quad x_2 \sim U(0,1)$$

$$B \rightarrow x_1 \sim U(0,1) \quad x_2 \sim \chi^2(1)$$

$$C \rightarrow x_1 \sim \chi^2(1) \quad x_2 \sim \chi^2(1)$$

The variable w is always generated as N(3,1). The second column shows the value of the parameters β_1 and β_2 in each experiment. The intercept is fixed at -2 for all simulations.

As in the previous alternatives of Case 1, columns IC and C2 contain the number of times in the 500 replications, that the criterion chose M1. We grouped AIC and SBIC into IC, because the results of both criteria were identical. From Section 3 we know that, within the theoretical framework, all the criteria behave well when the competing models are equivalent with identical densities, which is the situation of experiments 27 and 3. We can observe that their column IC and C2 values are close to 250, which means that the criteria behaves well, corroborating the theoretical results.

On the other hand, experiment 6 corresponds to equivalent models with non-identical densities, and the theoretical conclusion was that the IC criteria always behave well, while C_2 only in some cases. By observing columns IC and C2 for experiment 6, we can see that in this case both IC and C_2 perform well. Given this result for C_2, we analysed the characteristics of this experiment: the DGP variables both have Uniform distribution, and the weights of both variables (values of β_1 and β_2) are similar, and not too high; this implies that the differences between $\beta_1 x_1$ and $\beta_2 x_2$ are small. We think that these characteristics could cause C_2 to behave well.

Now we analysed the situation where one model is better than the other. The theoretical result in this case enabled us to confirm that the behaviour of the three criteria is adequate. Given the order of Table 9, we find the experiment with equivalent models in the middle of this table. Above them (the upper part of the table) we see the experiments where M1 is better, and below them (the lower part) those where M2 is the best model. According the theoretical results, in the upper end of Table 9 the values in columns IC and C2 should tend to 500, and in the lower end should tend to zero. The results obtained corroborate this behaviour in general terms, and additionally we can see that the values in the IC and C2 columns are closer to 500 or 0 as AM1 goes to 2000 or zero, respectively.

Table 9: *Behaviour of the selection criteria for all experiments in Alternative 1.3. N = 2000*

Table 9. Part I								**Table 9. Part II**							
Nº	β_1 , β_2	***x***	**absdel**	**AM1**	**absdifden**	**IC**	**C2**	**Nº**	β_1 , β_2	***x***	**absdel**	**AM1**	**absdifden|**	**IC**	**C2**
104	(3,-0.5)	C	0.4787	1868	0,3525	500	500	**76**	(4,1)	B	0.2595	1366	0,2633	500	500
34	(7,1)	A	0.3660	1861	0,3369	500	500	**119**	(2,1)	C	0.3901	1365	0,3080	500	500
17	(3,0.5)	A	0.1725	1846	0,1867	500	500	**30**	(1.5,1)	A	0.0391	1359	0,1224	499	499
16	(3,-0.5)	A	0.1533	1834	0,2205	500	500	**13**	(3,-2)	A	0.1006	1342	0,1495	500	500
33	(5,1)	A	0.2985	1827	0,3085	500	500	**107**	(3,1.5)	C	0.4368	1321	0,3329	500	500
92	(1,-0.5)	C	0.2342	1825	0,2024	500	500	**26**	(-1.5,1)	A	0.0099	1317	0,0217	446	438
103	(3,-1)	C	0.4607	1809	0,3167	500	500	**88**	(7,3)	B	0.2996	1292	0,2539	500	500
91	(1,-1)	C	0.2388	1807	0,2006	500	500	**42**	(4,3)	A	0.1629	1288	0,1902	500	500
90	(1,-1.5)	C	0.2354	1794	0,1924	500	500	**118**	(1.5,1)	C	0.3518	1268	0,2843	500	500
102	(3,-1.5)	C	0.4174	1779	0,2734	500	500	**59**	(3,-1)	B	0.1604	1252	0,1980	500	500
89	(1,-2)	C	0.2295	1779	0,1854	500	500	**132**	(7,3)	C	0.4774	1219	0,3404	500	500
101	(3,-2)	C	0.4190	1753	0,2640	500	500	**108**	(3,2)	C	0.4355	1188	0,3306	500	500
32	(4,1)	A	0.2424	1751	0,2853	500	500	**131**	(5,3)	C	0.4601	1156	0,3359	500	500
105	(3, 0.5)	C	0.4493	1691	0,2460	500	500	**62**	(3,1)	B	0.2238	1154	0,2484	67	244
18	(3,1)	A	0.1643	1691	0,3525	500	500	**94**	(1,1)	C	0.2914	1112	0,1969	441	451
15	(3,-1)	A	0.1308	1675	0,1955	500	500	**58**	(3,-1.5)	B	0.1698	1112	0,2484	352	494
78	(7,1)	B	0.3325	1641	0,2970	500	500	**130**	(4,3)	C	0.4481	1093	0,3319	500	500
122	(7,1)	C	0.5512	1606	0,3913	500	500	**87**	(5,3)	B	0.2980	1067	0,2586	447	496
44	(7,3)	A	0.2278	1582	0,2181	500	500	**48**	(1,-0..5)	B	0.0152	1035	0,0325	72	195
121	(5,1)	C	0.5037	1556	0,3697	500	500	**57**	(3,-2)	B	0.1749	1011	0,1956	41	322
19	(3,1.5)	A	0.1502	1540	0,2331	500	500	**27**	(-1,1)	A	0.0072	1003	0,0224	245	240
120	(4,1)	C	0.4737	1538	0,3553	500	500	**3**	(1,-1)	A	0.0072	1001	0,0227	250	245
31	(2 ,1)	A	0.0744	1524	0,1704	500	500	**6**	(1,1)	A	0.0185	1000	0,0796	249	252
77	(5,1)	B	0.2901	1516	0,2762	500	500	**63**	(3,1.5)	B	0.2533	995	0,2597	0	0
5	(1,0.5)	A	0.0128	1514	0,0513	499	498	**21**	(3,3)	A	0.1346	991	0,1867	240	223
93	(1,0.5)	C	0.2467	1510	0,2251	500	500	**109**	(3,3)	C	0.4404	980	0,3307	357	324
14	(3,-1.5)	A	0.1119	1504	0,1701	500	500	**86**	(4,3)	B	0.3031	932	0,2686	0	2
4	(1,-0.5)	A	0.0074	1498	0,0261	484	484	**110**	(3,4)	C	0.4487	891	0,3329	0	0
60	(3,-0.5)	B	0.1485	1492	0,2036	500	500	**64**	(3,2)	B	0.2781	877	0,2700	0	0
106	(3,1)	C	0.4388	1474	0,3372	500	500	**95**	(1,1.5)	C	0.3364	835	0,2771	0	0
25	(-2.1)	A	0.0127	1469	0,0218	486	483	**111**	(3,5)	C	0.4547	813	0,2318	0	0
43	(5,3)	A	0.1916	1438	0,2019	500	500	**75**	(2,1)	B	0.2024	813	0,3349	0	0
61	(3,0.5)	B	0.1813	1429	0,2357	500	500	**129**	(2,3)	C	0.4403	805	0,3339	0	0
20	(3,2)	A	0.1365	1393	0,2140	500	500	**47**	(1,-1)	B	0.0201	786	0,0341	2	22

Table 9. Part III

Nº	β_1 , β_2	*x*	absdel	AM1	absdifden	IC	C2
65	(3,3)	B	0,3185	764	0,2878	0	0
112	(3,7)	C	0,4572	719	0,3348	0	0
128	(1.5,3)	C	0,4352	708	0,3338	0	0
96	(1,2)	C	0,3691	705	0,2994	0	0
22	(3,4)	A	0,1648	701	0,1907	0	0
66	(3,4)	B	0,3424	686	0,2987	0	0
7	(1,1.5)	A	0,0391	685	0,1221	2	2
2	(1,-1.5)	A	0,0100	673	0,0222	37	41
46	(1,-1.5)	B	0,0229	668	0,0349	0	5
35	(-2,3)	A	0,0996	662	0,1476	0	0
41	(2,3)	A	0,1353	636	0,2126	0	0
67	(3,5)	B	0,3561	601	0,3046	0	0
45	(1,-2)	B	0,0249	589	0,0354	0	2
117	(0.5,1)	C	0,2344	583	0,2161	0	0
97	(1,3)	C	0,4175	576	0,3295	0	0
23	(3,5)	A	0,1965	537	0,2050	0	0
68	(3,7)	B	0,3752	518	0,3138	0	0
98	(1,4)	C	0,4556	513	0,3497	0	0
1	(1,-2)	A	0,0132	511	0,0226	5	6
29	(0.5,1)	A	0,0128	510	0,0510	1	1
28	(-0.5,1)	A	0,0073	505	0,0257	10	13
8	(1,2)	A	0,0739	504	0,1691	0	0
36	(-1.5,3)	A	0,1093	494	0,1673	0	0
74	(1.5,1)	B	0,2110	491	0,2313	0	0
99	(1,5)	C	0,4849	486	0,3643	0	0
40	(1.5,3)	A	0,1484	475	0,2302	0	0
85	(2,3)	B	0,3654	444	0,3217	0	0
100	(1,7)	C	0,5234	414	0,3828	0	0
49	(1,0.5)	B	0,0979	390	0,1357	0	0
24	(3,7)	A	0,2350	385	0,2244	0	0
127	(0.5,3)	C	0,4203	329	0,3424	0	0
9	(1,3)	A	0,1629	325	0,2433	0	0
84	(1.5,3)	B	0,4062	315	0,1921	0	0
37	(-1,3)	A	0,1291	315	0,3445	0	0

Table 9. Part IV

Nº	β_1 , β_2	*x*	absdel	AM1	absdifden	IC	C2
50	(1,1)	B	0,2295	261	0,2339	0	0
123	(-2,3)	C	0,4387	245	0,2922	0	0
124	(-1.5,3)	C	0,4285	232	0,2930	0	0
51	(1,1.5)	B	0,3140	226	0,2856	0	0
10	(1,4)	A	0,2403	224	0,2829	0	0
114	(-1.5,1)	C	0,2179	209	0,1678	0	0
125	(-1,3)	C	0,4377	208	0,3116	0	0
52	(1,2)	B	0,3704	204	0,3191	0	0
11	(1,5)	A	0,2965	195	0,3075	0	0
113	(-2,1)	C	0,2393	194	0,1811	0	0
69	(-2,1)	B	0,2227	190	0,1946	0	0
115	(-1,1)	C	0,2276	180	0,1843	0	0
53	(1,3)	B	0,4476	170	0,2173	0	0
39	(0.5,3)	A	0,1714	170	0,2460	0	0
38	(-0.5,3)	A	0,1518	170	0,3619	0	0
116	(-0.5,1)	C	0,2280	167	0,1974	0	0
54	(1,4)	B	0,4993	162	0,2031	0	0
70	(-1.5,1)	B	0,2286	162	0,3903	0	0
126	(-0.5,3)	C	0,4568	161	0,3427	0	0
79	(-2,3)	B	0,4713	157	0,3571	0	0
12	(1,7)	A	0,3670	137	0,3393	0	0
71	(-1,1)	B	0,2356	133	0,2116	0	0
56	(1,7)	B	0,5789	131	0,4345	0	0
55	(1,5)	B	0,5323	129	0,4094	0	0
80	(-1.5,3)	B	0,4772	128	0,3638	0	0
73	(0.5,1)	B	0,2435	104	0,2323	0	0
81	(-1,3)	B	0,4853	97	0,3692	0	0
72	(-0.5,1)	B	0,2422	84	0,2192	0	0
83	(0.5,3)	B	0,4788	80	0,3722	0	0
82	(-0.5,3)	B	0,4929	57	0,3732	0	0

Nevertheless, we observed some experiments with non-adequate behaviour. In the upper part of the table, when values of IC and C2 should tend to 500, experiments 62, 48 and 57 have values far less than 500. There are also large differences between the values of the two columns. In the lower part of the table (when M2 is better than M1), we find that experiments 21 and 109 have IC and C2 column values far from 0. In order to analyse what might be the cause of this atypical behaviour, we carried out a study of the measure *absdel*$_i$ = |E (ℓ^*_{1i}) - E (ℓ^*_{2i})|, for $i = 1, \cdots, 2000$, for these five experiments. We sequenced all observations from the lowest value to the highest value of *absdel*$_i$ and we found that, except in experiment 21, in the case of extreme values (the biggest or the smallest values) all were associated with the same sign of E (ℓ^*_{1i}) - E (ℓ^*_{2i}). In other words, all the observation corresponding to extreme values of *absdel*$_i$ tend to choose the same model. This fact can lead the average value *absdel* to be larger or smaller than they should be. In this case it could be that the values of IC and C2 were affected by these anomalous observations. To prove it, we have eliminated these extreme values, and we have again calculated every measure (*absdel*, AM1, IC and C2). The results are presented in Table 10.

Table 10: *Atypical experiments. Behaviour of the selection criteria.*

Experiment	β_1 , β_2	x	N	AM1	absdel	IC	C2
62	(3,1)	B	1800	1144	0,1517	500	500
48	(1,-0.5)	B	1750	1035	0,0097	377	499
57	(3,-2)	B	1800	1003	0,12	499	500
109	(3,3)	C	1925	909	0,4	0	0

As we can see in Table 10, when we eliminate the anomalous observations, the behaviour of the criteria becomes adequate. As we have previously mentioned, experiment 21 is the only one where this pattern is not maintained. In this case, extreme values of *absdel*$_i$ don't correspond to the same selection, and we haven't been able to find the cause of its atypical behaviour.

As a final comment, we observed a greatly reduced number of experiments with equivalent models. We understand that it is logical given that the experiments of alternative 1.3 correspond to pairs of models where the DGP is not nested in M1 or M2. Given that model M1 contains the variable x_1 and model M2 contains x_2, x_1 and x_2 being the only DGP variables, it is difficult to find cases where both the M1 and M2 models were equivalent.

We then re-executed the analysis using a sample size of 200. It is presented in Tables 11 through 14. As previously mentioned, we must be careful about the conclusions due to the use of asymptotic results to classify the experiments. The measures *absdel*, *absdifden* and AM1 can be affected, and it could lead to a less than appropriate classification.

We can see the experiments with the lowest value of *absdel* in Table 11. Only experiments 27 and 3 have a value of AM1 close to 100 (N/2). We can consider them the only experiments which could contain equivalent models. Table 12 presents the experiments with the values of AM1 closest 100 and it clearly corroborates that the models are only equivalent in experiments 27 and 3. Table 13 also shows the experiments with the lowest values of *absdifden*, and we can see that experiment 27 is in the first place. This experiment can be classified in the group corresponding to equivalent models with identical densities at the pseudo-true parameters. However, experiment 3 corresponds to the group of equivalent models with non-identical densities at the pseudo-true parameters.

Table 11: *Experiments with the lowest value of absdel*. N = 200.

	DGP					
Experiment	β'	x_1	x_2	**absdel**	AM1	absdifden
27	(-2,-1,1)	U	U	**0,00697284**	97	0,0214815
3	(-2,1,-1)	U	U	**0,00754769**	110	0,0228772
28	(-2,-0.5,1)	U	U	**0,00770398**	41	0,0253974
4	(-2,1,-0.5)	U	U	**0,00776623**	153	0,0260846
2	(-2,1,-1.5)	U	U	**0,00983076**	81	0,0221053

Table 12: *Experiments with values of AM1 around 100. N = 200.*

	DGP					
Experiment	β'	x_1	x_2	**absdel**	AM1	absdifden
6	(-2,1,1)	U	U	0,01879650	**92**	0,0807524
21	(-2,3,3)	U	U	0,16293388	**93**	0,2176746
64	(-2,3,2)	U	χ^2	0,29387544	**93**	0,2759691
110	(-2,3,4)	χ^2	χ^2	0,50047744	**93**	0,3505065
109	(-2,3,3)	χ^2	χ^2	0,45922314	**95**	0,3369146
27	(-2,-1,1)	U	U	0,00697284	**97**	0,0214815
57	(-2,3,-2)	U	χ^2	0,18829923	**99**	0,2035532
86	(-2,4,3)	U	χ^2	0,31628076	**99**	0,2746017
48	(-2,1,-0,5)	U	χ^2	0,01474792	**100**	0,033108
63	(-2,3,1,5)	U	χ^2	0,24814640	**100**	0,2549071
130	(-2,4,3)	χ^2	χ^2	0,46741192	**108**	0,3353497
58	(-2,3,-1,5)	U	χ^2	0,17616982	**109**	0,2021085
3	(-2,1,-1)	U	U	0,00754769	**110**	0,0228772

Table 13: *Experiments with the lowest values of absdifden. N = 200.*

	DGP					
Experiment	β'	x_1	x_2	absdel	AM1	**absdifden**
27	(-2,-1,1)	U	U	0,00697284	97	**0,0214815**
2	(-2,1,-1.5)	U	U	0,00754769	81	**0,0221053**
3	(-2,1,-1)	U	U	0,00770398	110	**0,0228772**
28	(-2,-0.5,1)	U	U	0,00776623	41	**0,0253974**
4	(-2,1,-0.5)	U	U	0,00983076	153	**0,0260846**

Similarly to sample size 2000, for each experiment Table 14 gives the number of times that model M1 was chosen. The experiments are sequenced from the highest to the lowest value of AM1. Columns IC and C2 should contain values close to 500 in the upper part, around 250 for the experiments with equivalent models (which have AM1 close to 100) and near zero for the lower part of the table. This behaviour is satisfied in almost all experiments; nevertheless, we found some atypical behaviour, which we are analyzing.

If we begin by focusing on the experiments placed between the two that contain equivalent models (experiments 3 and 27, placed in part II of Table 14), we can say that all results could be possible for IC and C2, given that the AM1 measure is close to N/2. Nevertheless we observe that in the case with AM1>N/2, the values in the IC and C2 columns tend to 500, while they tends to zero when AM1<N/2. From this, we can conclude that the observed behaviour is adequate.

Then we analyzed the upper part of Table 14 (above experiment 3), and we observed that only experiment 62 breaks the pattern of adequate values in the IC and C2 columns. On the other hand, in the lower part of Table 14 (below experiment 27) we find some anomalous results in experiments 109, 21 and 6. Following the same procedure as with N=2000, we tried to explain this atypical behaviour but, with a small sample size, we haven't found the cause of it. As we previously mentioned, we know that results and conclusions from a small sample size are not as clear and robust. In spite of that, we can affirm that the three criteria perform quite well in finite sample sizes.

Table 14: *Behaviour of the selection criteria in Alternative 1.3. N = 200*

Table 14. Part I								Table 14. Part II							
Nº	β_1 , β_2	x	absdel	AM1	absdifden	IC	C2	**Nº**	β_1 , β_2	x	absdel	AM1	absdifden|	IC	C2
89	(1,-2)	C	0.2797	186	0.1762	500	500	**118**	(1.5,1)	C	0.3202	133	0.1048	500	500
16	(3,-0.5)	A	0.1537	185	0.1249	500	500	**107**	(3,1.5)	C	0.4130	132	0.1456	500	500
17	(3,0.5)	A	0.1704	184	0.1085	500	500	**43**	(5,3)	A	0.2146	131	0.1494	495	494
90	(1,-1.5)	C	0.2558	184	0.1721	500	500	**88**	(7,3)	B	0.3157	131	0.2055	500	499
104	(3,-0.5)	C	0.4603	183	0.2235	500	500	**132**	(7,3)	C	0.4727	131	0.1821	500	500
103	(3,-1)	C	0.4798	183	0.1932	500	500	**20**	(3,2)	A	0.1453	128	0.0780	489	485
34	(7,1)	A	0.3535	181	0.2363	500	500	**59**	(3,-1)	B	0.1605	127	0.1353	434	458
105	(3,0.5)	C	0.4435	181	0.1952	500	500	**30**	(1.5,1)	A	0.0380	126	0.0816	411	412
92	(1,-0.5)	C	0.2377	179	0.1552	500	500	**42**	(4,3)	A	0.1889	123	0.1224	468	470
33	(5,1)	A	0.2947	175	0.1851	500	500	**62**	(3,1)	B	0.2021	121	0.0759	218	283
102	(3,-1.5)	C	0.4246	175	0.1510	500	500	**131**	(5,3)	C	0.4774	116	0.1742	498	498
32	(4,1)	A	0.2395	172	0.1450	500	500	**108**	(3,2)	C	0.4186	115	0.1465	500	499
91	(1,-1)	C	0.2406	172	0.1531	500	500	**94**	(1,1)	C	0.2477	114	0.0770	477	479
15	(3, -1)	A	0.1341	171	0.1354	500	500	**87**	(5,3)	B	0.3230	111	0.1831	389	454
122	(7,1)	C	0.5385	167	0.2184	500	500	**3**	(1,-1)	A	0.0075	110	0.0408	**258**	**272**
18	(3,1)	A	0.1612	164	0.0961	500	500	**58**	(3,-1.5)	B	0.1762	109	0.1481	257	334
121	(5,1)	C	0.5186	163	0.1999	500	500	**130**	(4,3)	C	0.4674	108	0.1693	494	491
120	(4,1)	C	0.4782	162	0.1841	500	500	**48**	(1,-0.5)	B	0.0147	100	0.0561	172	204
106	(3,1)	C	0.4223	161	0.3597	500	500	**63**	(3,1..5)	B	0.2481	100	0.1051	12	35
14	(3,-1.5)	A	0.1184	158	0.1388	500	498	**57**	(3,-2)	B	0.1883	99	0.1569	137	222
44	(7,3)	A	0.2410	158	0.1914	500	500	**86**	(4,3)	B	0.3163	99	0.1628	66	176
93	(1,0.5)	C	0.2292	157	0.1104	500	500	**27**	(-1,1)	A	0.0070	97	0.0389	**250**	**251**
61	(3,0.5)	B	0.1755	155	0.0709	499	499	**109**	(3,3)	C	0.4592	95	0.1672	397	381
101	(3,-2)	C	0.3448	155	0.1017	500	500	**21**	(3,3)	A	0.1629	93	0.0964	236	241
60	(3,-0.5)	B	0.1525	154	0.1257	500	500	**64**	(3,2)	B	0.2939	93	0.1279	1	5
4	(1,-0.5)	A	0.0078	153	0.0428	373	368	**110**	(3,4)	C	0.5005	93	0.1781	171	173
78	(7,1)	B	0.3355	153	0.2164	500	500	**6**	(1,1)	A	0.0188	92	0.0748	241	250
31	(2,1)	A	0.0721	150	0.0753	491	492	**111**	(3,5)	C	0.5105	88	0.1834	51	58
19	(3,1.5)	A	0.1515	148	0.0833	500	500	**65**	(3,3)	B	0.3294	83	0.1539	0	0
5	(1,0.5)	A	0.0124	146	0.0599	387	383	**47**	(1,-1)	B	0.0209	82	0.0625	77	115
13	(3,-2)	A	0.1090	146	0.1362	484	488	**113**	(-2,1)	C	0.1762	82	0.1116	0	0
77	(5,1)	B	0.2948	144	0.1694	500	500	**112**	(3,7)	C	0.4958	82	0.1936	4	5
119	(2,1)	C	0.3630	144	0.1291	500	500	**2**	(1,-1.5)	A	0.0098	81	0.0420	180	187
76	(4,1)	B	0.2520	143	0.1233	497	499	**95**	(1,1.5)	C	0.2899	81	0.0872	137	149

Table 14. Part III

Nº	β_1 , β_2	x	absdel	AM1	absdifden	IC	C2
129	(2,3)	C	0,4561	79	0,1671	81	78
22	(3,4)	A	0,1957	76	0,1257	36	40
75	(2,1)	B	0,1772	75	0,0699	1	1
128	(1.5,3)	C	0,4489	73	0,1611	2	2
66	(3,4)	B	0,3357	72	0,1736	0	0
114	(-1.5,1)	C	0,1620	70	0,1114	0	0
67	(3,5)	B	0,3480	70	0,1872	0	0
41	(2,3)	A	0,1536	68	0,0831	10	10
23	(3,5)	A	0,2232	68	0,1507	1	3
96	(1,2)	C	0,3368	68	0,1061	3	6
46	(1,-1.5)	B	0,0235	67	0,0655	56	79
7	(1,1.5)	A	0,0412	62	0,0852	68	71
98	(1,4)	C	0,5004	62	0,1664	0	0
117	(0.5,1)	C	0,1671	61	0,0873	2	4
99	(1,5)	C	0,5479	61	0,1753	0	0
68	(3,7)	B	0,3655	60	0,2018	0	0
45	(1,-2)	B	0,0249	59	0,0670	54	66
85	(2,3)	B	0,3811	58	0,1582	0	0
97	(1,3)	C	0,4210	58	0,1445	0	0
35	(-2,3)	A	0,0995	55	0,1357	18	22
8	(1,2)	A	0,0776	52	0,0799	7	6
100	(1,7)	C	0,5698	52	0,1898	0	0
40	(1.5,3)	A	0,1581	48	0,0883	0	0
29	(0.5,1)	A	0,0134	47	0,0629	91	93
49	(1,0.5)	B	0,0849	45	0,0871	0	0
24	(3,7)	A	0,2441	44	0,1858	0	0
74	(1.5,1)	B	0,1896	43	0,0932	0	0
36	(-1.5,3)	A	0,1139	42	0,1425	2	2
28	(-0.5,1)	A	0,0077	41	0,0422	146	145
84	(1.5,3)	B	0,4258	36	0,1730	0	0
9	(1,3)	A	0,1690	34	0,1018	0	0
115	(-1,1)	C	0,1359	33	0,1180	0	0
127	(0.5,3)	C	0,4091	28	0,1650	0	0
52	(1,2)	B	0,3832	26	0,1713	0	0

Table 14. Part IV

Nº	β_1 , β_2	x	absdel	AM1	absdifden	IC	C2
10	(1,4)	A	0,2477	25	0,1481	0	0
69	(-2,1)	B	0,1942	24	0,1695	0	0
50	(1,1)	B	0,2090	23	0,1207	0	0
11	(1,5)	A	0,3071	23	0,1859	0	0
37	(-1,3)	A	0,1372	22	0,1420	0	0
70	(-1.5,1)	B	0,2024	22	0,1721	0	0
116	(-0.5,1)	C	0,1680	21	0,1523	0	0
54	(1,4)	B	0,5059	21	0,2055	0	0
53	(1,3)	B	0,4759	19	0,1887	0	0
71	(-1,1)	B	0,2096	18	0,1727	0	0
79	(-2,3)	B	0,5005	18	0,2460	0	0
55	(1,5)	B	0,5281	18	0,2225	0	0
39	(0.5,3)	A	0,1774	16	0,1140	0	0
38	(-0.5,3)	A	0,1599	15	0,1322	0	0
51	(1,1.5)	B	0,3050	15	0,1488	0	0
12	(1,7)	A	0,3724	15	0,2320	0	0
123	(-2,3)	C	0,5429	15	0,1907	0	0
56	(1,7)	B	0,5717	15	0,2391	0	0
125	(-1,3)	C	0,4711	14	0,2393	0	0
80	(-1.5,3)	B	0,5085	14	0,2470	0	0
124	(-1.5,3)	C	0,5341	13	0,2152	0	0
72	(-0.5,1)	B	0,2185	10	0,1691	0	0
126	(-0.5,3)	C	0,4508	10	0,2322	0	0
73	(0.5,1)	B	0,2233	9	0,1460	0	0
81	(-1,3)	B	0,5161	9	0,2420	0	0
82	(-0.5,3)	B	0,5256	8	0,2317	0	0
83	(0.5,3)	B	0,5139	5	0,2044	0	0
130	(-0.5,1)	B	0,2422	84	0,1731	0	0
131	(0.5,3)	B	0,4788	80	0,2054	0	0
132	(-0.5,3)	B	0,4929	57	0,2276	0	0

MONTECARLO EXERCISE FOR CASE 2

$$M1: p_i = F(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i})$$
$$M2: p_i = F(\delta_0 + \delta_1 x_{1i} + \delta_2 z_i)$$

$$\text{PGD: } p_i = F\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}\right)$$

Similarly to Alternatives 1.1 and 1.2 of Case 1, we know which is the best model so we don't need to elaborate the complicated design of Alternative 1.3. Model M1 is the best, and it should be reflected in the results contained in the following tables.

From Tables 15 and 16, we can see that the theoretical results we obtained in Section 3 for Case 2 have been corroborated.

Table 15. *Behaviour of the selection criteria for Case 2. N = 200*

DGP			**Specified model**				
β'	x_1	x_2	w	**z**	**AIC**	**SBIC**	C_2
(-2,3,1)	U	U	U	N(3,1)	487	487	483
(-2,1,3)	U	U	U	N(3,1)	500	500	500
(-2,1,1)	U	U	U	N(3,1)	470	470	462
(-2,3,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	*dm*	U	N(3,1)	500	500	500
(-2,1,3)	U	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	U	*dm*	U	N(3,1)	499	499	499
(-2,3,1)	χ^2	*dm*	U	N(3,1)	495	495	495
(-2,1,3)	χ^2	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	*dm*	U	N(3,1)	499	499	499

Table 16. *Behaviour of the selection criteria for Case 2. N = 2000*

DGP			Specified model				
β'	x_1	x_2	w	z	AIC	SBIC	C_2
(-2,3,1)	U	U	U	N(3,1)	500	500	500
(-2,1,3)	U	U	U	N(3,1)	500	500	500
(-2,1,1)	U	U	U	N(3,1)	500	500	500
(-2,3,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,3)	U	χ^2	U	N(3,1)	500	500	500
(-2,1,1)	U	χ^2	U	N(3,1)	500	500	500
(-2,3,1)	U	*dm*	U	N(3,1)	500	500	500
(-2,1,3)	U	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	U	*dm*	U	N(3,1)	500	500	500
(-2,3,1)	χ^2	*dm*	U	N(3,1)	500	500	500
(-2,1,3)	χ^2	*dm*	U	N(3,1)	500	500	500
(-2,1,1)	χ^2	*dm*	U	N(3,1)	500	500	500

5. - CONCLUSIONS.

The study of some model selection criteria applied to overlapping binary models was the purpose of this work. We focused on the AIC, SBIC and C_2 criteria. The first two are recognised as Information Criteria (IC), while C_2 is based on the Mean Square Error of Prediction.

In this context, two binary models are overlapping if both are the same functional form (both probit or both logit), have some common explanatory variables and some specific variables. In this article we analyzed the methodology for selecting the best model among a group of binary overlapping models. We considered only two competing models, but the results obtained can be extended to selection among a larger number of models. This is one of the advantages of the selection criteria against the hypotheses testing procedures, because the latter only chooses between a pair of models.

One important aspect of our work is that one of the alternatives we considered is the situation where the true model (DGP) is not included in the two competing models.

We also analyzed the more common situation considering that one of the two models is correctly specified, that is to say, one of them belongs to the DGP family of models.

We carried out theoretical analysis to find out the behaviour of the three criteria. Then, if we know the DGP and the two competing models, we could find (theoretically) the relationships between the latter: i) if the pseudo-true parameter vector of one model converges to the true parameter vector, then this model is the closest to the DGP, and it is better than the other model; ii) if both pseudo-true parameter vectors converge to the true vector, both competing models are equivalent; iii) if both pseudo-true parameter vectors do not converge to the true vector, then any relationship can occur (equivalent or not), and it becomes the most interesting and complicated case to analyze.

Once we knew the theoretical relationship between the competing models, the objective of our study was to find the behaviour of the three selection criteria. To achieve that, we needed to answer the question: what is the requirement that a given criterion must satisfy to be considered as adequate? Specifically, if two models are equivalent, the probability limits of a given criterion must be the same in both models. However, if one of them is better, its corresponding limit must be lower than that of the other model. The theoretical analysis carried out confirms that all the criteria perform well in every situation. Only C_2 does not behave well sometimes in a specific alternative.

These theoretical results have been corroborated through a Monte Carlo experiment. The most complicated situation to simulate was, as we expected, that of two models whose pseudo-true parameter vectors don't converge to the true, so any relation between them can happen. In this case, we had to generate 132 different DGPs, leading to 132 different experiments. We want to point out that this is the situation where a small sample size affects more the results. Additionally, with these experiments we were able to characterise the specific situation previously mentioned, where C_2 may not perform well.

In general terms, we can say that these three criteria behave well in overlapping binary models: when the two competing model aren't the DGP, the criteria tend to choose the best of them, that is to say, the closest to the DGP. In the most commonly studied case where the DGP is included in one of the competing models, our conclusion is that the criteria also perform well, as we expected.

APPENDIX

Appendix 1: *Derivation of C_2 criterion*

Following Linhart and Zucchini (1986), the general expression of any criterion can be derived as:

$$\mathrm{R} = \Delta_N(\hat{\beta}) + \frac{tr\,\hat{\Omega}^{-1}\hat{\Sigma}}{N} \tag{A1.1}$$

where $\hat{\Omega}$ and $\hat{\Sigma}$ are consistent estimations of the Ω and Σ matrices, whose elements adopt the following form:

$$\Omega_{ij} = E_0\left[\frac{\partial^2\Delta(\beta^*)}{\partial\beta_i\partial\beta_j}\right]$$

$$\Sigma_{ij} = E_0\left[\left(\frac{\partial\Delta(\beta^*)}{\partial\beta_i}\right)\left(\frac{\partial\Delta(\beta^*)}{\partial\beta_j}\right)\right] \qquad i,j = 1,2,\ldots.k$$

Next, we derive the expression (A1.1)) for the discrepancy underlying the C_2 criterion, that is:

$$\Delta(\beta) = E_0(y_p - F(x_p'\beta))^2 \tag{A1.2}$$

where *p* denotes an out sample observation.

From (A1.2) the associated empirical discrepancy is given by:

$$\Delta_N(\beta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - F(x_i'\beta))^2 \tag{A1.3}$$

The first term of (A1.1) is obtained replacing β in (A1.3) with the minimum discrepancy estimator, which is calculated as:

$$\hat{\beta} = \arg\min_\beta \frac{1}{N}\sum_{i=1}^{N}(y_i - F(x_i'\beta))^2$$

Given the discrepancy (A1.2), in order to obtain the second term of the criterion we take into account that:

$$\Omega = E_0\left[\frac{\partial^2(y_p - F(x_p'\beta^*))^2}{\partial\beta\partial\beta'}\right] \tag{A1.4}$$

$$\Sigma = E_0\left[\frac{\partial(y_p - F(x_p'\beta^*))^2}{\partial\beta}\right]\left[\frac{\partial(y_p - F(x_p'\beta^*))^2}{\partial\beta'}\right] \tag{A1.5}$$

To develop these expressions, we use the following results:

$$\frac{\partial(y_p - F(x_p'\beta))^2}{\partial\beta} = -2(y_p - F(x_p'\beta))\ f(x_p'\beta)\ x_p \tag{A1.6}$$

$$\frac{\partial^2 (y_p - F(x_p'\beta))^2}{\partial\beta\partial\beta'} = -2\left[-f^2(x_p'\beta)\right] x_p x_p' - 2(y_p - F(x_p'\beta)) f'(x_p'\beta) x_p x_p' \tag{A1.7}$$

Substituting (A1.7) in (A1.4) and evaluating β at β^*, we obtain:

$$\Omega = E_0(2 f^2(x_p'\beta^*) x_p x_p') = 2 f^2(x_p'\beta^*) x_p x_p'$$

Since under correct specification, it holds that $E_0(y_p - F(x_p'\beta^*)) = 0$, the estimation for the Ω matrix is:

$$\hat{\Omega} = 2 f^2(x_p'\hat{\beta}) x_p x_p' \tag{A1.8}$$

Substituting (A1.6) in (A1.5) and evaluating β at β^*, results that:

$$\Sigma = 4E_0\{(y_p - F(x_p'\beta^*))^2\}\ f^2(x_p'\beta^*) x_p x_p'$$

and its estimation is:

$$\hat{\Sigma} = 4\{\frac{1}{N}\sum_{i=1}^{N} (y_i - F(x_i'\hat{\beta}))^2\}\ f^2(x_p'\hat{\beta}) x_p x_p' \tag{A1.9}$$

From (A1.8) and (A1.9) we obtain:

$$\hat{\Omega}^{-1}\hat{\Sigma} = 2\ [\ f^2(x_p'\hat{\beta}) x_p x_p'\ ]^{-1}\ [\ f^2(x_p'\hat{\beta}) x_p x_p'\ ]\{\frac{1}{N}\sum_{i=1}^{N} (y_i - F(x_i'\hat{\beta}))^2\}$$

$$= 2\ I_{\mathrm{k}} \frac{1}{N}\sum_{i=1}^{N} (y_i - F(x_i'\hat{\beta}))^2$$

and:

$$tr\ (\hat{\Omega}^{-1}\hat{\Sigma}) = \frac{2k}{N} SSD \tag{A1.10}$$

with $SSD = \sum_{i=1}^{N} (y_i - F(x_i'\hat{\beta}))^2$

Taking into account (A1.3), but using $\hat{\beta}$ instead of β, together with (A1.10), we obtain the C_2 criterion expressed in (9):

$$C_2 = \frac{1}{N}\sum_{i=1}^{N} (y_i - F(x_i'\hat{\beta}))^2 + \frac{2k}{N^2}\sum_{i=1}^{N} (y_i - F(x_i'\hat{\beta}))^2 = \frac{SSD}{N}\left(1 + \frac{2k}{N}\right)$$

Appendix 2: *Study of the performance of the criterion* C_2 *when* $F_{0i} \in$ *(0, 0.5) or (0.5,1).*

As we can see in (52), the superiority condition of M1 against M2 is written as follows:

$$\left(\frac{F_{1i}}{F_{2i}}\right)^{F_{0i}} > \left(\frac{1-F_{2i}}{1-F_{1i}}\right)^{1-F_{0i}} \tag{A2.1}$$

2.1. $F_{0i} \in (0, 0.5)$.

In this range the following may occur:

2.1.a. $F_{1i} > F_{2i}$ (A2.2)

2.1.b. $F_{1i} < F_{2i}$ (A2.3)

2.1.a. From (A2.2), we see that $\frac{F_{1i}}{F_{2i}} > 1$ and $\frac{1-F_{2i}}{1-F_{1i}} > 1$, and both terms decrease with the exponents F_{0i} and $1-F_{0i}$ respectively, but the decrease will be larger for the one with the smallest exponent. Given that we can't clearly establish such a relationship, we analyse both situations separately.

i) $\frac{F_{1i}}{F_{2i}} > \frac{1-F_{2i}}{1-F_{1i}}$ (A2.4)

Here, (A2.1) can be only satisfied if F_{0i} is not too small, that is to say, $F_{0i} \cong 0.5$. Continuing (A2.4), we have:

$$F_{2i}^2 - F_{1i}^2 > F_{2i} - F_{1i} \Rightarrow (F_{2i} + F_{1i}) \cdot (F_{2i} - F_{1i}) > (F_{2i} - F_{1i}) \tag{A2.5}$$

and given that $F_{1i} > F_{2i}$ we obtain $F_{1i} + F_{2i} < 1$.

ii) $\frac{F_{1i}}{F_{2i}} < \frac{1-F_{2i}}{1-F_{1i}}$ (A2.6)

Given that $F_{0i} <$ (1- F_{0i}), we can conclude that under (A2.1), condition (A2.6) will never be satisfied.

2.1.b. From (A2.3) we see that $\frac{F_{1i}}{F_{2i}} < 1$ and $\frac{1-F_{2i}}{1-F_{1i}} < 1$, and both terms increase with the corresponding exponents; the term with the lowest exponent increases more than the other term. As in the previous case, we have two situations:

i) $\frac{F_{1i}}{F_{2i}} > \frac{1-F_{2i}}{1-F_{1i}}$

Similarly to (A2.5) this case implies that $(F_{1i} + F_{2i}) > 1$. Given the relationship between the exponents ($F_{0i} <$ (1- F_{0i})), the superiority condition (A2.1) will be satisfied.

ii) $\frac{F_{1i}}{F_{2i}} < \frac{1-F_{2i}}{1-F_{1i}}$

From this expression we find that $(F_{1i}+F_{2i}) < 1$. In this case, (A2.1) will be satisfied only if F_{0i} is small enough ($F_{0i} \cong 0$).

In summary, when $F_{0i} \in (0, 0.5)$, the situations where the superiority condition is satisfied are the following:

$A.\quad F_{1i} < F_{2i} \quad y \quad F_{1i}+F_{2i} > 1$
$B.\quad F_{1i} > F_{2i} \quad y \quad F_{1i}+F_{2i} < 1 \quad y \quad F_{0i} \cong 0.5$
$C.\quad F_{1i} < F_{2i} \quad y \quad F_{1i}+F_{2i} < 1 \quad y \quad F_{0i} \cong 0$

In order to verify if these three situations lead to the C_2 criteria behaving well we analyse:

$$plim\,[C_2\text{ (M1) - }C_2\text{ (M2)}] = \frac{\sum_{i=1}^{N}\left[(F_{1i}-F_{2i})\,[F_{1i}+F_{2i}-2F_{0i}]\right]}{N} \qquad \text{(A2.7)}$$

This criterion will behave adequately if:

$$plim\,[C_2\text{ (M1) - }C_2\text{ (M2)}] < 0 \qquad \text{(A2.8)}$$

In Situation *A*, it is evident that the inequality (A2.8) holds regardless of the value of F_{0i} in the analyzed range. For *B*, we can see that if F_{0i} is not small, it will very often happen that $F_{1i}+F_{2i} < 2F_{0i}$ and then, (A2.8) will hold. Nevertheless, we want to point out that we assume that every relationship we are analyzing is satisfied on average, and that it can occur that the corresponding inequality does not hold for some *i* elements. For *C*, it is very possible that $F_{1i}+F_{2i} - 2F_{0i} > 1$ because $F_{0i} \cong 0$ and then (A2.8) is satisfied.

2.2. $F_{0i} \in (0.5, 1)$. We again distinguish between two possible cases:

2.2.a $F_{1i} > F_{2i}$ (A2.9)

2.2.b $F_{1i} < F_{2i}$ (A2.10)

2.2.a. From (A2.9) we have $\frac{F_{1i}}{F_{2i}} > 1$ and $\frac{1-F_{2i}}{1-F_{1i}} > 1$ and, both terms will decrease with the exponent; the term with the lowest exponent will decrease more than the other. Analogously to case 2.1.a., it can occur that:

i) $\frac{F_{1i}}{F_{2i}} > \frac{1-F_{2i}}{1-F_{1i}}$

This inequality leads to $(F_{1i} + F_{2i})$ < 1, and the superiority condition (A2.1) will always hold, given the relationship between the exponents $F_{0i} > (1 - F_{0i})$.

ii) $\dfrac{F_{1i}}{F_{2i}} < \dfrac{1 - F_{2i}}{1 - F_{1i}}$

This expression implies that $(F_{1i} + F_{2i})$ > 1. To arrive at the condition that (A2.1) is satisfied it is required that $F_{0i} \cong 1$.

<u>2.2.b.</u> From (A2.10), $\dfrac{F_{1i}}{F_{2i}}$ < 1 and $\dfrac{1 - F_{2i}}{1 - F_{1i}}$ < 1 so both will increase with their corresponding exponents; the one with the lowest exponent will be the term which will increase the most. We again analyse the two possible situations:

i) $\dfrac{F_{1i}}{F_{2i}} > \dfrac{1 - F_{2i}}{1 - F_{1i}}$

We have $(F_{1i} + F_{2i})$ > 1 and the superiority condition (A2.1) will be satisfied if F_{0i} is not too large, that is to say, $F_{0i} \cong 0.5$.

ii) $\dfrac{F_{1i}}{F_{2i}} < \dfrac{1 - F_{2i}}{1 - F_{1i}}$

This leads to $(F_{1i} + F_{2i})$ < 1, but it will not be satisfied under (A2.1)

In summary, the cases with the condition of superiority when $F_{0i} \in$ (0,5, 1) are:

$$\begin{array}{llllll} D. & F_{1i} > F_{2i} & y & F_{1i} + F_{2i} < 1 & & \\ E. & F_{1i} > F_{2i} & y & F_{1i} + F_{2i} > 1 & y & F_{0i} \cong 1 \\ F. & F_{1i} < F_{2i} & y & F_{1i} + F_{2i} > 1 & y & F_{0i} \cong 0.5 \end{array}$$

Now we want to verify that C_2 performs well in the three situations D, E and F. To do that, we use expressions (A2.7) and (A2.8). In D, given that $F_{0i} \in$ (0.5, 1) inequality (A2.8) is clearly satisfied. In *E*, inequality $F_{1i} + F_{2i} < 2F_{0i}$ will probably hold if F_{0i} is close to 1 and then, (A2.8) will be satisfied. In *F*, it is easy that $2F_{0i} \cong 1$ because $F_{0i} \cong 0.5$, meaning that $F_{1i} + F_{2i} - 2F_{0i}$ will be positive and (A2.8) verified.

Appendix 3: *Derivation of the relationship between the true parameter vector and the pseudo-true parameter vectors*

We assume two binary models:

MA: $p_i = F\,(x_i'\beta)$

MB: $p_i = F\,(z_i'\alpha)$

Given that our models are overlapping, the cumulative distribution function (F) is the same in both specifications, and the matrices X and Z are not nested.
Without loss of generality, we assume MA is the DGP.
To find the relationship between the true parameter vector (β) and the pseudo-true parameter vector (α^*) we use the first order condition associated to the pseudo-true parameters:

$$E_0\left[\frac{\partial \ell_B}{\partial \alpha}\right] = 0$$

where E_0 denotes the expectation under MA and ℓ_B is the log-likelihood function associated with MB. This condition, defined in White (1982), results in:

$$\sum_{i=1}^{N}\frac{F(x_i'\beta)-F(z_i'\alpha^*)}{F(z_i'\alpha^*)[1-F(z_i'\alpha^*)\ ]}f(z_i'\alpha^*)z_i = 0$$

$$\Rightarrow \sum_{i=1}^{N}F(x_i'\beta)\,A(z_i'\alpha^*)\,z_i = \sum_{i=1}^{N}F(z_i'\alpha^*)\,A(z_i'\alpha^*)\,z_i \qquad \text{(A3.1)}$$

where $A(\bullet) = \dfrac{f(\bullet)}{F(\bullet)\left[1-F(\bullet)\right]}$.

From (A3.1), we use the Taylor extensions $F(z_i'\alpha^*)\,A(z_i'\alpha^*)$ and $A(z_i'\alpha^*)$ **,** both around β**.** To consider a general framework, we define m_i' a row vector of variables which includes all the non-repeated variables of x_i' and z_i'**.** Additionally, we define $\beta^\bullet$ and $\alpha^{*\bullet}$, which are extended β and α, with zero elements in the places corresponding to variables not included in x_i' and z_i'**,** respectively. The mentioned Taylor extensions are:

$$F(z_i'\alpha^*)\,A(z_i'\alpha^*) = F(m_i'\alpha^{*\bullet})\,A(m_i'\alpha^{*\bullet})$$

$$\cong F(m_i'\beta^\bullet)\,A(m_i'\beta^\bullet) + [\ f(m_i'\beta^\bullet)\,A(m_i'\beta^\bullet) + F(m_i'\beta^\bullet)\,A'(m_i'\beta^\bullet)\ ]\ m_i'\,(\alpha^{*\bullet}-\beta^\bullet) \qquad \text{(A3.2)}$$

$$A(z_i'\alpha^*) = A(m_i'\alpha^{*\bullet}) \cong A(m_i'\beta^\bullet) + A'(m_i'\beta^\bullet)\ \ m_i'\,(\alpha^{*\bullet}-\beta^\bullet) \qquad \text{(A3.3)}$$

Replacing (A3.2) and (A3.3) into (A3.1) results:

$$\sum_{i=1}^{N}F(m_i'\beta^\bullet)\,A(m_i'\beta^\bullet)\,z_i + \sum_{i=1}^{N}F(m_i'\beta^\bullet)\,A'(m_i'\beta^\bullet)\ \ m_i'\,(\alpha^{*\bullet}-\beta^\bullet)\ \ z_i$$

$$= \sum_{i=1}^{N} F(m_i'\beta^\bullet)\, A(m_i'\beta^\bullet)\, z_i \; + \; \sum_{i=1}^{N} \; H(m_i'\beta^\bullet)\, m_i'(\alpha^{*\bullet} - \beta^\bullet) \;\; z_i \qquad \text{(A3.4)}$$

Being:

$$H(m_i'\beta^\bullet) \; = \; f(m_i'\beta^\bullet)\, A(m_i'\beta^\bullet) \;\; + \sum_{i=1}^{N} F(m_i'\beta^\bullet)\, A'(m_i'\beta^\bullet) \qquad \text{(A3.5)}$$

Simplifying (A3.4)[3] we obtain:

$$\sum_{i=1}^{N} F(m_i'\beta^\bullet)\, A'(m_i'\beta^\bullet) \;\; m_i'(\alpha^{*\bullet} - \beta^\bullet) \;\; z_i \; = \; \sum_{i=1}^{N} H(m_i'\beta^\bullet) \;\; m_i'(\alpha^{*\bullet} - \beta^\bullet) \;\; z_i \qquad \text{(A3.6)}$$

and we also replace expression (A3.5) of $H(m_i'\beta^\bullet)$, resulting in:

$$\sum_{i=1}^{N} \; f(m_i'\beta^\bullet)\, A(m_i'\beta^\bullet) \;\; m_i'(\alpha^{*\bullet} - \beta^\bullet) \;\; z_i \; = 0$$

$$\Rightarrow \sum_{i=1}^{N} \; f(m_i'\beta^\bullet)\, A(m_i'\beta^\bullet) \;\; m_i' \;\; \alpha^{*\bullet} \; z_i \; = \sum_{i=1}^{N} \; f(m_i'\beta^\bullet)\, A(m_i'\beta^\bullet) \;\; m_i' \;\; \beta^\bullet \; z_i \qquad \text{(A3.7)}$$

We know that $m_i' \;\; \alpha^{*\bullet} = \; z_i' \; \alpha^*$ and $m_i' \;\; \beta^\bullet \; = x_i' \;\; \beta$, so (A3.7) can be equivalently expressed as:

$$\sum_{i=1}^{N} [\, f(x_i'\beta)\, A(x_i'\beta) \;\; (z_i' \; \alpha^*)] \;\; z_i \; = \; \sum_{i=1}^{N} [\, f(x_i'\beta)\, A(x_i'\beta) \;\; (x_i' \;\; \beta)] \;\; z_i$$

or in matrix notation:

$$Z'\,V\,Z \;\; \alpha^* \; = \; Z'\,V\,X \;\; \beta \qquad \text{(A3.8)}$$

where Z and X are $(N{\cdot}k_2)$ and $(N{\cdot}k_1)$ matrices, α^* and β are $(k_2 \cdot 1)$ and $(k_1 \cdot 1)$ vectors , respectively, and V is a diagonal matrix with elements $f(x_i'\beta)\, A(x_i'\beta)$, $i = 1, \cdots, N$.

From (A3.8) we obtain an expression of α^* as a function of β:

$$\alpha^* \; = \; (Z'V\,Z)^{-1} \;\; Z' \;\; V X \;\; \beta$$

[3] Specifically, result (A3.4) would be an approximation against equality, because the residual terms of both Taylor expansions are different; nevertheless, such residual terms can be considered as negligible for a fixed sample size. But if N tends to infinite (which is our usual framework), we must be sure these residual terms are zero, which is demonstrated by using the Central Limit Theorem.

REFERENCES

Akaike, H. (1973). Information Theory and a Extension of the Likelihood Ratio Principle. In: Proceedings of the Second International Symposium of Information Theory. Budapest: Akademiai Kiado, pp. 189-204.

Amemiya, T (1985). Advanced Econometrics. Cambridge: Harvard University Press,.

Aparicio, T., Villanua, I. (2007). Some Selection Criteria for Nested Binary Choice Models: A Comparative Study. Computational Statistics 22: 635-660.

Davidson, J. (2000). Econometric Theory. Oxford: Blackwell.

Gourieroux, C., Monfort, A., Renault, E., Trognon, A. (1987). Simulated Residuals. Journal of Econometrics 34: 201-252.

Kim, H., Cavanaugh, J.E. (2005). Model Selection Criteria based on Kullback Information Measures for Nonlinear Regression. Journal of Statistical Planning and Inference 134: 332-349.

Linhart, H., Zucchini, W. (1986). Model Selection. New York: John Wiley and Sons.

Monfardini, C. (2003). An Illustration of Cox's Non-Nested Testing Procedure for Logit and Probit Models. Computational Statistics and Data Analysis 42: 425-444.

Newey, W.K., McFadden, D. (1994). Large sample estimation and hypothesis testing. In Handbook of Econometrics 4. Amsterdam: Elsevier Science, pp. 2113-2241.

Pesaran, M.H., Pesaran, B. (1993). A Simulation Approach to the Problem of Computing Cox's Statistics for Testing Nonnested Models. Journal of Econometrics 57: 377-392.

Santos Silva, J.M.C. (2001). A Score Test for Non-Nested Hypotheses with Applications to Discrete Data Models. Journal of Applied Econometrics 16: 577-597.

Schwarz, G. (1978). Estimating the Dimension of a Model". The Annals of Statistics 6 (2): 461-464.

Van Der Hoeven, N. (2005). The probability to select the correct model using likelihood-ratio based criteria in choosing between two nested models of which the more extended one is true. Journal of Statistical Planning and Inference 135: 477-486.

Vuong, Q.H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses". Econometrica, 57(2): 307-333.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. Econometrica, 50(1): 1-25.

White, H. (2001). Asymptotic Theory for Econometricians. San Diego: Academic Press.

DOCUMENTOS DE TRABAJO

Facultad de Economía y Empresa

Universidad de Zaragoza

Depósito Legal Z-1411-2010. ISSN 2171-6668

2002-01: "Evolution of Spanish Urban Structure During the Twentieth Century". Luis Lanaspa, Fernando Pueyo y Fernando Sanz. Department of Economic Analysis, University of Zaragoza.

2002-02: "Una Nueva Perspectiva en la Medición del Capital Humano". Gregorio Giménez y Blanca Simón. Departamento de Estructura, Historia Económica y Economía Pública, Universidad de Zaragoza.

2002-03: "A Practical Evaluation of Employee Productivity Using a Professional Data Base". Raquel Ortega. Department of Business, University of Zaragoza.

2002-04: "La Información Financiera de las Entidades No Lucrativas: Una Perspectiva Internacional". Isabel Brusca y Caridad Martí. Departamento de Contabilidad y Finanzas, Universidad de Zaragoza.

2003-01: "Las Opciones Reales y su Influencia en la Valoración de Empresas". Manuel Espitia y Gema Pastor. Departamento de Economía y Dirección de Empresas, Universidad de Zaragoza.

2003-02: "The Valuation of Earnings Components by the Capital Markets. An International Comparison". Susana Callao, Beatriz Cuellar, José Ignacio Jarne and José Antonio Laínez. Department of Accounting and Finance, University of Zaragoza.

2003-03: "Selection of the Informative Base in ARMA-GARCH Models". Laura Muñoz, Pilar Olave and Manuel Salvador. Department of Statistics Methods, University of Zaragoza.

2003-04: "Structural Change and Productive Blocks in the Spanish Economy: An Imput-Output Analysis for 1980-1994". Julio Sánchez Chóliz and Rosa Duarte. Department of Economic Analysis, University of Zaragoza.

2003-05: "Automatic Monitoring and Intervention in Linear Gaussian State-Space Models: A Bayesian Approach". Manuel Salvador and Pilar Gargallo. Department of Statistics Methods, University of Zaragoza.

2003-06: "An Application of the Data Envelopment Analysis Methodology in the Performance Assessment of the Zaragoza University Departments". Emilio Martín. Department of Accounting and Finance, University of Zaragoza.

2003-07: "Harmonisation at the European Union: a difficult but needed task". Ana Yetano Sánchez. Department of Accounting and Finance, University of Zaragoza.

2003-08: "The investment activity of spanish firms with tangible and intangible assets". Manuel Espitia and Gema Pastor. Department of Business, University of Zaragoza.

2004-01: "Persistencia en la performance de los fondos de inversión españoles de renta variable nacional (1994-2002)". Luis Ferruz y María S. Vargas. Departamento de Contabilidad y Finanzas, Universidad de Zaragoza.

2004-02: “Calidad institucional y factores político-culturales: un panorama internacional por niveles de renta”. José Aixalá, Gema Fabro y Blanca Simón. Departamento de Estructura, Historia Económica y Economía Pública, Universidad de Zaragoza.

2004-03: “La utilización de las nuevas tecnologías en la contratación pública”. José Mª Gimeno Feliú. Departamento de Derecho Público, Universidad de Zaragoza.

2004-04: “Valoración económica y financiera de los trasvases previstos en el Plan Hidrológico Nacional español”. Pedro Arrojo Agudo. Departamento de Análisis Económico, Universidad de Zaragoza. Laura Sánchez Gallardo. Fundación Nueva Cultura del Agua.

2004-05: “Impacto de las tecnologías de la información en la productividad de las empresas españolas”. Carmen Galve Gorriz y Ana Gargallo Castel. Departamento de Economía y Dirección de Empresas. Universidad de Zaragoza.

2004-06: “National and International Income Dispersión and Aggregate Expenditures”. Carmen Fillat. Department of Applied Economics and Economic History, University of Zaragoza. Joseph Francois. Tinbergen Institute Rotterdam and Center for Economic Policy Resarch-CEPR.

2004-07: “Targeted Advertising with Vertically Differentiated Products”. Lola Esteban and José M. Hernández. Department of Economic Analysis. University of Zaragoza.

2004-08: “Returns to education and to experience within the EU: are there differences between wage earners and the self-employed?”. Inmaculada García Mainar. Department of Economic Analysis. University of Zaragoza. Víctor M. Montuenga Gómez. Department of Business. University of La Rioja

2005-01: “E-government and the transformation of public administrations in EU countries: Beyond NPM or just a second wave of reforms?”. Lourdes Torres, Vicente Pina and Sonia Royo. Department of Accounting and Finance.University of Zaragoza

2005-02: “Externalidades tecnológicas internacionales y productividad de la manufactura: un análisis sectorial”. Carmen López Pueyo, Jaime Sanau y Sara Barcenilla. Departamento de Economía Aplicada. Universidad de Zaragoza.

2005-03: “Detecting Determinism Using Recurrence Quantification Analysis: Three Test Procedures”. María Teresa Aparicio, Eduardo Fernández Pozo and Dulce Saura. Department of Economic Analysis. University of Zaragoza.

2005-04: “Evaluating Organizational Design Through Efficiency Values: An Application To The Spanish First Division Soccer Teams”. Manuel Espitia Escuer and Lucía Isabel García Cebrián. Department of Business. University of Zaragoza.

2005-05: “From Locational Fundamentals to Increasing Returns: The Spatial Concentration of Population in Spain, 1787-2000”. María Isabel Ayuda. Department of Economic Analysis. University of Zaragoza. Fernando Collantes and Vicente Pinilla. Department of Applied Economics and Economic History. University of Zaragoza.

2005-06: “Model selection strategies in a spatial context”. Jesús Mur and Ana Angulo. Department of Economic Analysis. University of Zaragoza.

2005-07: “Conciertos educativos y selección académica y social del alumnado”. María Jesús Mancebón Torrubia. Departamento de Estructura e Historia Económica y Economía Pública.

Universidad de Zaragoza. Domingo Pérez Ximénez de Embún. Departamento de Análisis Económico. Universidad de Zaragoza.

2005-08: "Product differentiation in a mixed duopoly". Agustín Gil. Department of Economic Analysis. University of Zaragoza.

2005-09: "Migration dynamics, growth and convergence". Gemma Larramona and Marcos Sanso. Department of Economic Analysis. University of Zaragoza.

2005-10: "Endogenous longevity, biological deterioration and economic growth". Marcos Sanso and Rosa María Aísa. Department of Economic Analysis. University of Zaragoza.

2006-01: "Good or bad? - The influence of FDI on output growth. An industry-level analysis". Carmen Fillat Castejón. Department of Applied Economics and Economic History. University of Zaragoza. Julia Woerz. The Vienna Institute for International Economic Studies and Tinbergen Institute, Erasmus University Rotterdam.

2006-02: "Performance and capital structure of privatized firms in the European Union". Patricia Bachiller y Mª José Arcas. Departamento de Contabilidad y Finanzas. Universidad de Zaragoza.

2006-03: "Factors explaining the rating of Microfinance Institutions". Begoña Gutiérrez Nieto and Carlos Serrano Cinca. Department of Accounting and Finance. University of Saragossa, Spain.

2006-04: "Libertad económica y convergencia en argentina: 1875-2000". Isabel Sanz Villarroya. Departamento de Estructura, Historia Económica y Economía Pública. Universidad de Zaragoza. Leandro Prados de la Escosura. Departamento de Hª e Instituciones Ec. Universidad Carlos III de Madrid.

2006-05: "How Satisfied are Spouses with their Leisure Time? Evidence from Europe*". Inmaculada García, José Alberto Molina y María Navarro. University of Zaragoza.

2006-06: "Una estimación macroeconómica de los determinantes salariales en España (1980-2000)". José Aixalá Pastó y Carmen Pelet Redón. Departamento de Estructura, Historia Económica y Economía Pública. Universidad de Zaragoza.

2006-07: "Causes of World Trade Growth in Agricultural and Food Products, 1951 – 2000". Raúl Serrano and Vicente Pinilla. Department of Applied Economics and Economic History, University of Zaragoza, Gran Via 4, 50005 Zaragoza (Spain).

2006-08: "Prioritisation of patients on waiting lists: a community workshop approach". Angelina Lázaro Alquézar. Facultad de Derecho, Facultad de Económicas. University of Zaragoza. Zaragoza, Spain. Begoña Álvarez-Farizo. C.I.T.A.- Unidad de Economía. Zaragoza, Spain

2007-01: "Deteminantes del comportamiento variado del consumidor en el escenario de Compra". Carmén Berné Manero y Noemí Martínez Caraballo. Departamento de Economía y Dirección de Empresas. Universidad de Zaragoza.

2007-02: "Alternative measures for trade restrictiveness. A gravity approach". Carmen Fillat & Eva Pardos. University of Zaragoza.

2007-03: "Entrepreneurship, Management Services and Economic Growth". Vicente Salas Fumás & J. Javier Sánchez Asín. Departamento de Economía y Dirección de Empresas. University of Zaragoza.

2007-04: "Equality versus Equity based pay systems and their effects on rational altruism motivation in teams: Wicked masked altruism". Javier García Bernal & Marisa Ramírez Alerón. University of Zaragoza.

2007-05: "Macroeconomic outcomes and the relative position of Argentina´s Economy: 1875-2000". Isabel Sanz Villarroya. University of Zaragoza.

2008-01: "Vertical product differentiation with subcontracting". Joaquín Andaluz Funcia. University of Zaragoza.

2008-02: "The motherwood wage penalty in a mediterranean country: The case of Spain" Jose Alberto Molina Chueca & Victor Manuel Montuenga Gómez. University of Zaragoza.

2008-03: "Factors influencing e-disclosure in local public administrations". Carlos Serrano Cinca, Mar Rueda Tomás & Pilar Portillo Tarragona. Departamento de Contabilidad y Finanzas. Universidad de Zaragoza.

2008-04: "La evaluación de la producción científica: hacia un factor de impacto neutral". José María Gómez-Sancho y María Jesús Mancebón-Torrubia. Universidad de Zaragoza.

2008-05: "The single monetary policy and domestic macro-fundamentals: Evidence from Spain". Michael G. Arghyrou, Cardiff Business School and Maria Dolores Gadea, University of Zaragoza.

2008-06: **"**Trade through fdi: investing in services". Carmen Fillat-Castejón, University of Zaragoza, Spain; Joseph F. Francois. University of Linz, Austria; and CEPR, London & Julia Woerz, The Vienna Institute for International Economic Studies, Austria.

2008-07: "Teoría de crecimiento semi-endógeno vs Teoría de crecimiento completamente endógeno: una valoración sectorial". Sara Barcenilla Visús, Carmen López Pueyo, Jaime Sanaú. Universidad de Zaragoza.

2008-08: "Beating fiscal dominance. The case of spain, 1874-1998". M. D. Gadea, M. Sabaté & R. Escario. University of Zaragoza.

2009-01: "Detecting Intentional Herding: What lies beneath intraday data in the Spanish stock market" Blasco, Natividad, Ferreruela, Sandra (Department of Accounting and Finance. University of Zaragoza. Spain); Corredor, Pilar (Department of Business Administration. Public University of Navarre, Spain).

2009-02: "What is driving the increasing presence of citizen participation initiatives?". Ana Yetano, Sonia Royo & Basilio Acerete. Departamento de Contabilidad y Finanzas. Universidad de Zaragoza.

2009-03: "Estilos de vida y "reflexividad" en el estudio del consumo: algunas propuestas". Pablo García Ruiz. Departamento de Psicología y Sociología. Universidad de Zaragoza.

2009-04: "Sources of Productivity Growth and Convergence in ICT Industries: An Intertemporal Non-parametric Frontier Approach". Carmen López-Pueyo and Mª Jesús Mancebón Torrubia. Universidad de Zaragoza.

2009-05: **"**Análisis de los efectos medioambientales en una economía regional: una aplicación para la economía aragonesa". Mónica Flores García y Alfredo J. Mainar Causapé. Departamento de Economía y Dirección de Empresas. Universidad de Zaragoza.

2009-06: "The relationship between trade openness and public expenditure. The Spanish case, 1960-2000". Mª Dolores Gadea, Marcela Sabate y Estela Saenz. Department of Applied Economics. School of Economics. University of Economics.

2009-07: "Government solvency or just pseudo-sustainability? A long-run multicointegration approach for Spain". Regina Escario, María Dolores Gadea, Marcela Sabaté. Applied Economics Department. University of Zaragoza.

2010-01: "Una nueva aproximación a la medición de la producción científica en revistas JCR y su aplicación a las universidades públicas españolas". José María Gómez-Sancho, María Jesús Mancebón Torrubia. Universidad de Zaragoza

2010-02: "Unemployment and Time Use: Evidence from the Spanish Time Use Survey". José Ignacio Gimenez-Nadal, University of Zaragoza, José Alberto Molina, University of Zaragoza and IZA, Raquel Ortega, University of Zaragoza.

2011-1: "Universidad y Desarrollo sostenible. Análisis de la rendición de cuentas de las universidades del G9 desde un enfoque de responsabilidad social". Dr. José Mariano Moneva y Dr. Emilio Martín Vallespín, Universidad de Zaragoza.

2011-2: "Análisis Municipal de los Determinantes de la Deforestación en Bolivia." Javier Aliaga Lordeman, Horacio Villegas Quino, Daniel Leguía (Instituto de Investigaciones Socio-Económicas. Universidad Católica Boliviana), y Jesús Mur (Departamento de Análisis Económico. Universidad de Zaragoza)

2011-3: "Imitations, economic activity and welfare". Gregorio Giménez. Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza.

2012-1: "Selection Criteria for Overlapping Binary Models". M. T Aparicio and I. Villanúa. Department of Economic Analysis, Faculty of Economics, University of Zaragoza