Parametric and semiparametric specification tests for binary choice models: a comparative simulation study*

Isabel Proença and J.M.C. Santos Silva
ISEG/Universidade Técnica de Lisboa

April 28, 2001

Abstract

Although semiparametric alternatives are available, parametric binary choice models are widely used in practice, in spite of their sensitivity to misspecification. Here we present the results of a simulation study on the finite sample performance of parametric and semiparametric specification tests for this kind of models. The results obtained indicate that the computationally demanding semiparametric tests do not generally outperform the simpler score tests against parametric alternatives.

Word count: 1691.

Keywords: HH-test; Logit; Score tests.

JEL Classification codes: C12, C14, C25, C52.

*Address for correspondence: R. do Quelhas 6, 1200 Lisboa, Portugal. E-mail: isabelp@iseg.utl.pt (Proença) and jmcss@iseg.utl.pt (Santos Silva). Fax: 351 213922781. We thank J.M. Andrade e Silva for valuable comments on a previous version of this paper. The usual disclaimer applies. Financial support from Fundação para a Ciência e a Tecnologia, programme Praxis XXI, is gratefully acknowledged.

1 INTRODUCTION

Despite the availability of semiparametric estimators, parametric binary choice models are still widely used in applied econometrics. Examples of application of these models can be found in areas such as travel demand, credit scoring, firm failures and participation in the job market. See the classical paper by Amemiya (1981) and the recent monograph by Gourieroux (2000) for many other examples.

Because parametric binary choice models are generally not robust to misspecification, it is important to test for departures from the null hypothesis. Davidson and MacKinnon (1984) introduced a simple way to construct regression-based score tests of the null against parametric alternatives. More recently, Horowitz and Härdle (1994) have developed a new approach to the construction of specification tests for binary choice models. If under the null the model is of the single index type, then it can be tested against a model with the same index but with an unspecified link function, using simple semiparametric techniques. The attractive feature of the test against the semiparametric alternative is that it may have power against a wider range of departures from the null hypothesis than the simpler score tests. However, the semiparametric test is computationally much more demanding.

In this paper we present the results of a simulation study comparing the finite sample performance of the semiparametric test with that of several specification tests based on the more traditional score test approach. Throughout, the popular logit model is taken as the null hypothesis.

2 MODEL AND TESTS

Consider a binary random variable Y_i such that

$$P\left(Y_i = 1|x_i\right) = E\left(Y_i|x_i\right) = F\left(x_i^T\beta\right) \quad i = 1, \cdots, n$$

where x_i is a vector of covariates, β is a conformable vector of parameters and $F(\cdot)$ is a non-decreasing continuous link function such that $0 \leq F(-\infty) < F(\infty) \leq 1$.

Given $F(\cdot)$, $\hat{\beta}$, the maximum likelihood estimates of β, can be obtained as described in any standard textbook (see for example Davidson and MacKinnon, 1993). In what follows, we concentrate on specification tests to detect departures from the null hypothesis $H_0 : E(Y_i|x_i) = F_0\left(x_i^T\beta\right) = \left\{1 + \exp\left(-x_i^T\beta\right)\right\}^{-1}$.

2.1 Score tests

A simple way to check the correct specification of $E(Y_i|x_i)$ is to test the null against a more general model that has $F_0\left(x_i^T\beta\right)$ as a special case. In this paper we consider score tests against the three following generalizations of the logit model:

a) $E(Y_i|x_i) = \left\{1 + \exp\left(-\,\theta\left(\rho x_i^T\beta\right)/\,\rho\right)\right\}^{-1}$, where ρ is a parameter and $\theta(\cdot)$ is a function such that $\theta(0) = 0$, $\theta'(0) = 1$ and $\theta''(0) \neq 0$ (see Davidson and MacKinnon, 1993, p. 527);

b) $E(Y_i|x_i) = \int \left\{1 + \exp\left(-x_i^T\beta - \varepsilon_i\right)\right\}^{-1} g(\varepsilon|x_i)\, d\varepsilon$, where ε_i is a random variable with conditional density $g(\varepsilon|x_i)$ and $E(\varepsilon_i|x_i) = 0$ and $E(\varepsilon_i^2|x_i) = \omega$;

c) $E(Y_i|x_i) = \left\{1 + \exp\left(-x_i^T\beta/\,\sigma\left(z_i^T\delta\right)\right)\right\}^{-1}$, where $\sigma(\cdot) > 0$ is a function such that $\sigma(0) = 1$ and $\sigma'(0) \neq 0$, and z_i denotes a vector containing the i-th observation of all explanatory variables except the intercept.

Tests against these alternatives can be performed as score tests for the omission of certain test variables, as described in Davidson and MacKinnon (1984) and Orme (1988). Specifically, the test for $\rho = 0$ in a) is a RESET-type test of the kind advocated by Pagan and Vella (1989) that checks for the omission of the additional regressor defined by $\left(x_i^T\hat{\beta}\right)^2$; the test against b) can be interpreted as an information matrix test (Chesher, 1984) for which the test variable is $1 - 2F_0\left(x_i^T\hat{\beta}\right)$; finally the score test against c) is a test against an heteroskedastic logit of the type discussed by Davidson and MacKinnon (1984) in which the test regressors are given by $\left(x_i^T\hat{\beta}\right) z_i$. Under the null, the RESET-type and the information matrix test statistics follow an asymptotic

$\chi^2_{(1)}$ distribution, while the statistic for the test against the heteroskedastic logit has an asymptotic $\chi^2_{(k-1)}$ distribution, where k is the dimension of x_i.

2.2 The semiparametric test

The motivation behind the semiparametric test is that the adequacy of the parametric link can be assessed by evaluating the weighted difference between the parametric estimate, $F_0\left(x_i^T\hat{\beta}\right)$, and the nonparametric estimate of $F(\cdot)$, given $x_i^T\hat{\beta}$. To perform the semiparametric test in the context of binary choice models, the test procedure of Horowitz and Härdle (1994), also known as the HH–test, can be used. The HH-statistic conveniently standardized is asymptotically distributed as a standard normal variate. However, simulations in Proença (1995) revealed that the distribution of the HH statistic in finite samples is bandwidth dependent and has a negative bias, which affects the performance of the test. These problems led Proença and Ritter (1994) to propose a modified test, the MHH-test, which is less bandwidth dependent than the HH-test, and has the same asymptotic distribution (see Härdle, Mammen and Proença, 2001). Moreover, they have deduced analytical corrections to the bias and variance that make the asymptotic critical values more accurate in finite samples. This corrections make the calculations of the test statistic more burdensome, but the development of computational capacities makes it possible to apply the MHH-test to problems with reasonably large data sets.

The MHH–statistic is based on $T_n = \sqrt{h}\hat{r}^T W\hat{r}$, where $\hat{r}$ is a vector of residuals with typical element $\hat{r}_i = Y_i - F_0\left(x_i^T\hat{\beta}\right)$ and W is a $n \times n$ matrix, the smoothing matrix, corresponding to leave-one-out kernel smoothing of the residuals with bandwidth h. The elements of W are defined by

$$w_{ij} = \begin{cases} \frac{K\left[\left(x_j^T\hat{\beta}-x_i^T\hat{\beta}\right)/h\right]}{\sum_{j\neq i} K\left[\left(x_j^T\hat{\beta}-x_i^T\hat{\beta}\right)/h\right]} u\left(x_i^T\hat{\beta}\right) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

where $u\left(x_i^T\hat{\beta}\right)$ is a weight function that equals 1 for the 90% central values of the ordered $x_i^T\hat{\beta}$, being zero otherwise.[1]

Let X denote the design matrix obtained by stacking the vectors x_i^T, and define $D = (I-H)^T W (I-H)$, where $H = VX\left(X^TVX\right)^{-1}X^T$, and V is a diagonal matrix with diagonal elements defined as $v_i = F_0\left(x_i^T\hat{\beta}\right)\left[1 - F_0\left(x_i^T\hat{\beta}\right)\right]$. Denoting by d_i the i-th diagonal element of D, the estimates for the bias correction and corrected variance, when the null is the logit, are given respectively by $bc = \sqrt{h}\ \sum_{i=1}^n d_i v_i$ and $\hat{\sigma}_n^2 = 2\,\mathrm{tr}\,(DVDV) + \sum_{i=1}^n d_i^2\left[v_i - 6v_i^2\right]$. Finally, the test statistic can be computed as $T_n^{MHH} = (T_n - bc)/\sqrt{\hat{\sigma}_n^2}$, which under the null has an asymptotic standard normal distribution.

3 SIMULATION RESULTS

In this section we present the results of a simulation study comparing the empirical size and power of the score and semiparametric specification tests described in section 2. For all experiments the index function was assumed to be $1 + x_{i1} + x_{i2}$, where x_{i1} comes from a standard normal and x_{i2} from a Bernoulli distribution with parameter 0.75. The null hypothesis is the logit model. Besides $F_0\left(x_i^T\beta\right)$, three other link functions were used to generate the data:

$$F_1\left(x_i^T\beta\right) = 1 - \exp\left\{-\exp\left(x_i^T\beta\right)\right\}, \quad (1)$$

$$F_2\left(x_i^T\beta\right) = \left\{1 + \exp(-x_i^T\beta)\right\}^{-1} - \frac{5(x_i^T\beta)}{9}\phi\left(\frac{(x_i^T\beta)}{1.5}\right), \quad (2)$$

$$F_3\left(x_i^T\beta\right) = \left\{1 + \exp\left(-x_i^T\beta/\sqrt{0.5 + 0.5x_{i1}^2}\right)\right\}^{-1}, \quad (3)$$

where $\phi(\cdot)$ denotes the standard normal density. Model (1) is the classic complementary log-log. Model (2) is a logit model perturbed by a bump. The bump was chosen so that the conditional expectation of Y_i is non-decreasing with $x_i^T\beta$. In this model, $P(Y_i = 1|x_i)$ is almost flat for values of the index near zero, indicating indecision of the agents when there is no clear superiority of one alternative. Finally, model (3) is a

[1] Different cut-off points can be used for these weights (see Horowitz and Härdle, 1994).

logit with heteroskedasticity. When data are generated by $F_0\left(x_i^T\beta\right)$ the empirical size of the tests is assessed, while for data generated by the other models the percentage of rejections gives the empirical power of the tests.

The response was generated from a Bernoulli distribution with probability of success given by the four models described above. The variables were generated using the same random seed for each model and the regressors were newly drawn in each replication. The sample sizes considered were 250, 500 and 750. For each experiment 5000 replications were performed.

The score tests were computed using the artificial regression approach suggested by Davidson and MacKinnon (1984). In the semiparametric test the bandwidth for kernel estimation is of the form $h = cn^{-0.2}$, and the value of c was chosen using a two-step procedure. In a first step, a graphical inspection of the kernel regression estimate was used to define tentative values of the smoothing parameter, leading to the following choices for c: 0.6, 1.5 and 2.4. The larger value of h slightly over-smooths the data, whereas the smaller h somewhat under-smooths it. In a second step, pilot simulation studies were used to evaluate the behaviour of the test using these choices and it was found that over-smoothing was clearly desirable. After some experimentation with different degrees of over-smoothing, the values of c chosen were 2.4, 4.5 and 6.6. The semiparametric test was performed using a one-sided critical region of the type $\left\{T_n^{MHH} > z_\alpha\right\}$, where z_α is the $(1-\alpha)$ percentile of the standard normal (see Horowitz and Härdle, 1994). Table 1 gives the percentage of rejections of the null at the nominal 5% level. Percentage of rejections at the nominal 10% level are available from the authors upon request.

Under the null, all parametric tests perform reasonably well even in samples of size 250. The performance of the semiparametric test is somewhat erratic for the smaller sample size, but it improves rapidly with the number of observations. Under the alternative, all parametric tests have reasonable power against a wide range of alternatives. Naturally, their performance depends on the particular nature of the misspecification. Once again, the behaviour of the semiparametric test clearly depends on the smoothing parameter. In these experiments, the semiparametric test is never clearly superior to the score tests, and most of the times it is out-performed by them. It is important to notice that these results were obtained with values of the bandwidth chosen so that a good compromise between size and power could be obtained. In an applied setting, some exploratory simulations can also be used to choose the bandwidth parameter. Although that will certainly lead to more reliable inference, it implies a heavy computational burden.

Table 1: Percentage of rejections at the nominal 5% level

Sample size	True model	Score tests against: a)	b)	c)	Semiparametric test $c = 2.4$	$c = 4.5$	$c = 6.6$
250	$F_0\left(x_i^T\beta\right)$	0.0394	0.0502	0.0466	0.0686	0.0398	0.0120
	$F_1\left(x_i^T\beta\right)$	0.5396	0.8706	0.8276	0.4422	0.6586	0.7440
	$F_2\left(x_i^T\beta\right)$	0.2504	0.2308	0.2328	0.2606	0.1364	0.0422
	$F_3\left(x_i^T\beta\right)$	0.3710	0.2104	0.3002	0.2572	0.2210	0.1350
500	$F_0\left(x_i^T\beta\right)$	0.0414	0.0458	0.0466	0.0644	0.0504	0.0246
	$F_1\left(x_i^T\beta\right)$	0.7408	0.9952	0.9854	0.7174	0.9396	0.9896
	$F_2\left(x_i^T\beta\right)$	0.4896	0.4842	0.4440	0.4862	0.4552	0.2814
	$F_3\left(x_i^T\beta\right)$	0.5426	0.3186	0.4220	0.4402	0.4600	0.3512
750	$F_0\left(x_i^T\beta\right)$	0.0488	0.0464	0.0506	0.0644	0.0636	0.0454
	$F_1\left(x_i^T\beta\right)$	0.9314	1.0000	0.9996	0.8646	0.9914	1.0000
	$F_2\left(x_i^T\beta\right)$	0.6814	0.6806	0.6344	0.6572	0.6878	0.5668
	$F_3\left(x_i^T\beta\right)$	0.6776	0.4130	0.5454	0.5680	0.6228	0.5582

4 CONCLUSIONS

In the cases considered here, there is no clear superiority of one test procedure over its competitors. Moreover, although constructed with a specific alternative in mind, the score tests studied here are powerful against a wide range of alternatives and are not systematically outperformed by the semiparametric test. In view of these results, the computational burden of the semiparametric test is hard to justify, at least for routine use.

REFERENCES

Amemiya, T. (1981). "Qualitative Response Models: A Survey". *Journal of Economic Literature*, 19, 1483-1536.

Chesher, A. D. (1984). "Testing for Neglected Heterogeneity". *Econometrica* 52, 865-872.

Davidson, R. and MacKinnon, J. G. (1984). "Convenient Specification Tests for Logit and Probit Models". *Journal of Econometrics*, 25, 241-262.

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics.* Oxford University Press, Oxford.

Gourieroux, C. (2000). *The Econometrics of Qualitative Variables*, Cambridge University Press, Cambridge.

Härdle, W.; Mammen, E. and Proença, I. (2001)."A Bootstrap Test for Single Index Models", *Statistics*, to appear.

Horowitz, J. L. and Härdle, W. (1994). "Testing a Parametric Model Against a Semiparametric Alternative". *Econometric Theory*, 10, 821-848.

Orme, C. (1988). *Specification Tests for Binary Data Models.* Discussion Paper 65, Department of Economics, University of Nottingham.

Pagan, A. R. and Vella, F. (1989). "Diagnostic Tests for Models Based on Unit Record Data: A Survey". *Journal of Applied Econometrics*, 4, S29-S59.

Proença, I. (1995). *Testing the Link Specification in Binary Choice Models: A Semiparametric Approach.* Faculté des Sciences Economiques, Sociales et Politiques, Nouvelle Série-N^{o}259, CIACO Louvain-la-Neuve.

Proença, I. and Ritter, C. (1994). *Semiparametric Testing of the Link Function in Models for Binary Outcomes.* Discussion Paper 9415, Institut de Statistique, Université Catholique de Louvain.