Economics Bulletin

Volume 32, Issue 1

A supremum-type RESET test for binary choice models

Esmeralda Ramalho
Department of Economics and CEFAGE, Universidade de Evora

Joaquim Ramalho
Departament of Economics and CEFAGE, Universidade de Evora

Jose M.R. Murteira
Faculdade de Economia, Universidade de Coimbra and CEMAPRE

Abstract

This note introduces a supremum-type RESET statistic for testing the specification of binary choice regression models. A Monte Carlo simulation study reveals very promising results for the proposed statistic.

The authors thank the editor and the referees for helpful comments. Financial support from Fundação para a Ciência e a Tecnologia, program FEDER/POCI 2010, is gratefully acknowledged.
Citation: Esmeralda Ramalho and Joaquim Ramalho and Jose M.R. Murteira, (2012) "A supremum-type RESET test for binary choice models", *Economics Bulletin*, Vol. 32 No. 1 pp. 905-912.
Contact: Esmeralda Ramalho - ela@uevora.pt, Joaquim Ramalho - jsr@uevora.pt, Jose M.R. Murteira - jmurt@fe.uc.pt.
Submitted: January 09, 2012. **Published:** March 19, 2012.

1. Introduction

The RESET is the most widely used general specification test in the context of linear (Ramsey 1969) and nonlinear (Pagan and Vella 1989) single index regression models, consisting of a simple test for the significance of some fitted powers of the index function. This paper focus on the application of the RESET test in the binary regression framework, for which Ramalho and Ramalho (2012) have recently found that test versions using only one or two fitted powers are clearly the best RESET performers. However, while both RESET versions exhibit a similar behaviour in terms of size, their power performance is sometimes quite distinct. In fact, depending on the type and degree of misspecification considered, it may make a huge difference to use one version or the other. This compromises the usefulness of the RESET test in applied work, since, on the one hand, in general it is not possible to know *a priori* which RESET version is better and, on the other hand, using both RESET versions may lead to contradictory conclusions too often.

In order to circumvent the mentioned drawback of RESET tests, this note proposes a supremum-type RESET statistic. Supremum test statistics are commonly used in inference problems involving a nuisance parameter that is present only under the alternative hypothesis; see *inter alia* Andrews and Ploberger (1994) and Hansen (1996). In this note, it is argued that the choice of the number of fitted powers to include in the RESET test may be seen as an analogous problem to that of the choice of an arbitrary value for the nuisance parameter. The supremum-type RESET statistic is defined as the RESET version that exhibits the lowest p-value and the bootstrap is employed to approximate its empirical distribution. The finite sample properties of the new statistic are examined through a Monte Carlo analysis.

2. The supremum-type RESET test

Consider a sample of $i = 1, ..., N$ individuals and let $y = \{0, 1\}$ be the response variable of interest and x a vector of exogenous variables. The conditional expected value of y given x is defined as

$$E\left(y|x, \theta\right) = G\left(x\theta\right), \tag{1}$$

where $G\left(\cdot\right)$ is a cumulative density function and θ is the vector of parameters of interest. Typically, θ is estimated by (Bernoulli-based) maximum likelihood.

To investigate whether $G\left(x\theta\right)$ is a correct specification for $E\left(y|x,\theta\right)$, the RESET test may be employed. This test relies on the idea that any index model of the form $E\left(y|x\right) = F\left(x\theta\right)$ can be arbitrarily approximated by $G\left[x\theta + \sum_{j=1}^{J} \gamma_j \left(x\theta\right)^{j+1}\right]$ for J large enough. Therefore, testing the hypothesis H_0: $E\left(y|x\right) = G\left(x\theta\right)$ is equivalent to test for H_0: $\gamma = 0$ in the augmented model $E\left(y|x, z\right) = G\left(x\theta + z\gamma\right)$, where $z = \left[\left(x\hat{\theta}\right)^2, ..., \left(x\hat{\theta}\right)^{J+1}\right]$ and J is the dimension of z. According to the number of test variables included in z, different is the variant of the RESET test, denoted in this paper by RESETJ. Following Ramalho and Ramalho (2012), this paper considers only two RESET statistics: RESET1 and RESET2.

Given that RESET1 and RESET2 may have very different power properties in finite samples and that none of those RESET versions displays the best performance in all cases, combining somehow RESET1 and RESET2 into a single statistic may yield to a test that is sensitive to a wider variety of model misspecifications. In econometrics, the issue of combining different test versions into a single statistic is frequently addressed when a nuisance parameter is present only under the alternative hypothesis. In such a case, although the simple replacement of the nuisance parameter by any admissible value produces a valid test, it is much more common to compute a single test statistic that summarizes, according to a suitable criterion, the results calculated over a range of values of the unidentified nuisance parameter. The computation of the latter statistic requires in general that the bootstrap or similar procedures be used to obtain critical values, but avoids the choice of a particular value for the unidentified parameter that may sacrifice the power of the test. Clearly, this is a problem similar to that considered in this paper, since defining previously the number of test variables to compute the RESET test may compromise its small sample power performance. This paper focusses on the construction of a supremum version of the RESET test (Sup-RESET), but other criteria like simple or weighted averages could also be used; see Andrews and Ploberger (1994).

In general, computing a supremum statistic in the conventional nuisance parameter case involves the calculation of the supremum of the test statistics obtained for each value of a grid defined over many different values of the nuisance parameter. However, this procedure is valid only when all test variants share the same asymptotic distribution, which is not the case of RESET tests. Indeed, each RESET version assesses the significance of a different number of test variables and, hence, has a chi-square distribution with different degrees of freedom. Therefore, in the present framework, the Sup-RESET statistic is computed using the following procedures:

(i) Estimate model (1) in order to obtain estimates $\hat{\theta}$ for the parameters of interest.

(ii) Construct $\left(x\hat{\theta}\right)^2$ and $\left(x\hat{\theta}\right)^3$ and compute RESET1 and RESET2 tests using Wald, LR, LM or any other approach valid for testing the omission of covariates in nonlinear models;

(iii) Calculate the p-values associated to each RESET version;

(iv) Denote the lowest p-value by p^* and define as the Sup-RESET statistic the RESET version that yielded p^*.

As the Sup-RESET statistic does not have a known asymptotic null distribution, the bootstrap may be employed to approximate its empirical distribution in the following way:

(v) Generate B bootstrap samples by drawing with replacement the explanatory variables from the original sample and using (1) and $\hat{\theta}$ to generate the dependent variable;

(vi) For each bootstrap sample, use steps (i)-(iv) to compute the corresponding Sup-RESET statistic and save the associated p-value;

(vii) Compute the α quantile of the empirical distribution of the B bootstrap p-values, where α is the desired level of significance of the test, and denote it by p^{boot};

The null hypothesis of correct specification of $E\left(y|x,\theta\right)$ is rejected in case $p^* < p^{boot}$.

3. A Monte Carlo simulation study

In order to examine the finite sample properties of the Sup-RESET test, we conducted a small Monte Carlo simulation study based on the experimental design of Ramalho and Ramalho (2012). In all experiments we generated 2500 Monte Carlo samples of size 500. In most cases, we assumed a linear index with two covariates, $x\theta = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, with x_1 generated either as a standard normal or a displaced exponential variate and x_2 as a Bernoulli variate with mean 2/3. In the construction of the Sup-RESET statistic, we considered only LM versions of RESET1 and RESET2 and used 399 bootstrap replications. The empirical size and power of the Sup-RESET test is compared to that of RESET1 and RESET2, for which both asymptotic and bootstrap-based results are reported.

The size of RESET tests may be sensitive to the structural model that underlies the data and to the percentage of zeros/ones observed. Therefore, the results of Table 1 were obtained from data generated according to three models (cauchit, probit and loglog), with $(\theta_1, \theta_2) = (1, 1)$ and $\theta_0 = \{0, -2, 2\}$. All tests based on the use of bootstrap critical values exhibit a suitable size performance, since their actual sizes are not significantly different from the nominal size at a 5% level in all cases, unlike the RESET tests based on asymptotic theory.

The power of the Sup-RESET test is expected to depend on the type and degree of misspecification generated. We performed 28 experiments, described in Table 2, which concern the following issues:

(i) Misspecification of the link function $G(\cdot)$. In this case, the data was generated using a specification different from the one being tested.

(ii) Misspecification of the index function. We considered the omission of the covariate x_3 (associated to the parameter θ_3), which was generated as a displaced exponential variate with variance one, the omission of the square of x_1 (associated to the parameter θ_4), and the presence of heteroskedasticity determined by the skedastic function $s(x_1, \gamma) = e^{2\gamma x_1}$. The degree of misspecification increases as θ_3 and γ increase and θ_4 decreases.

(iii) Misspecification due to sampling issues: covariate measurement error, response misclassification and endogenous sampling. In order to control for the mechanism that governs those deviations, the functional form μ that describes the data is written as a function of $G(x\theta)$ and one or two additional parameters that define the misspecification mechanism: the variance of the measurement error, σ^2, the probability of observing 1 (0) when the actual response is 0 (1), δ_1 (δ_0), and the proportion of individuals for which $y = 1$ in the sample and in the population, H and Q. The degree of misspecification increases when σ^2, δ_1 and/or δ_0, and the difference between H and Q increase.

Figure 1 illustrates the power results of the Sup-RESET test along with those of bootstrap versions of RESET1 and RESET2. Results are reported only for experiments involving the probit model but similar findings were obtained for the other models. From Figure 1, it is clear that the power of Sup-RESET is not in general as large as the power of the best RESET version, be it RESET1 or RESET2, but is much closer to it than to the power of the least powerful RESET variant. For example, in experiment 3 the power of Sup-RESET is 83.3% while that of RESET1 and RESET2 are respectively 85.7% and 37.3%. A similar scenario occurs when RESET2 is more powerful than RESET1, see for instance experiments 1, 19, 20, 23 and 24, where the gains in power that result from using Sup-RESET instead of RESET2 range from 34.5% to 68.9%. Therefore, the Sup-RESET test has the very attractive property of avoiding the cases of low power that RESET1 and RESET2 sometimes exhibit.

4. Concluding remarks

This note proposes a new RESET test for binary choice models that presents a more reliable behaviour in finite samples than the conventional versions of the test. The implementation of the new test is very simple, since only involves the calculation of two conventional RESET statistics and the use of a bootstrap procedure that is easily implemented in most econometric software packages. Moreover, the test also presents the advantage of being straightforwardly extensible to test any other type of single index regression model and to incorporate other RESET versions.[1]

References

Andrews, D.W.K. and W. Ploberger (1994) "Optimal tests when a nuisance parameter is present only under the alternative" *Econometrica* **62(6)**, 1383-1414.

Hansen, B.E. (1996) "Inference when a nuisance parameter is not identified under the null hypothesis" *Econometrica* **64(2)**, 413-430.

Pagan, A. and F. Vella (1989) "Diagnostic tests for models based on individual data: a survey" *Journal of Applied Econometrics* **4**, S29-S59.

Ramalho, E.A. and J.J.S. Ramalho (2012) "Alternative versions of the RESET test for binary response index models: a comparative study" *Oxford Bulletin of Economics and Statistics* **74(1)**, 107-130.

Ramsey, J.B. (1969) "Tests for specification errors in classical linear least-squares regression analysis" *Journal of the Royal Statistical Society B* **31(2)**, 350-371.

[1] In the binary regression framework, some preliminary experiments revealed that the inclusion of additional RESET versions in the computation of Sup-RESET is not advantageous, producing systematically less powerful statistics.

Table 1: Empirical size

	Crit.		H_0: Cauchit			H_0: Probit			H_0: Loglog		
	Value	$\theta:$	$(0,1,1)$	$(-2,1,1)$	$(2,1,1)$	$(0,1,1)$	$(-2,1,1)$	$(2,1,1)$	$(0,1,1)$	$(-2,1,1)$	$(2,1,1)$
RESET1	asym.		4.0*	4.5	4.8	4.5	4.6	3.4*	4.8	4.0*	4.8
RESET2	asym.		4.5	4.5	4.7	4.4	4.3	3.4*	4.4	3.2*	4.9
RESET1	boot.		4.3	4.8	4.9	4.9	5.2	5.2	5.4	5.3	5.3
RESET2	boot.		4.4	4.8	4.8	4.9	5.2	5.1	5.1	4.7	5.0
Sup-RESET	boot.		4.4	4.8	4.9	4.8	5.2	5.2	5.1	5.0	4.9

Notes: $N = 500$; 2500 Monte Carlo replications; 399 bootstrap replications; the values starred are significantly different from the nominal size at the 5% level.

Table 2: Empirical power: experimental designs

Misspecification		Description	Experiment
Link	True model/null model:		
	probit/chauchit	$\theta = \{(0,1,1),(-2,1,1),(2,1,1)\}$	1,2,3
	probit/loglog		4,5,6
	cauchit/probit		7,8,9
	loglog/probit		10,11,12
Index	Omission of an uncorrelated covariate	$\theta_3 = \{1.25, 2.50\}$	13,14
	Omission of a quadratic term	$\theta_4 = \{-0.1, -0.2\}$	15,16
	Heteroskedasticity	$\gamma = \{0.15, 0.30\}$	17,18
Observation problems	Covariate measurement error	$\mu = G(x^*\theta)\left[1+\sigma^2 m(y,x^*)\right] + o(\sigma^2)$,* $\sigma^2 = \{0.5, 1\}$	19,20
	Response misclassification	$\mu = \delta_1 + (1-\delta_0-\delta_1)\, G(x\theta)$, $(\delta_0,\delta_1) = \{(0.125,0),(0.250,0),(0.0625,0.0625),(0.125,0.125)\}$	21,22,23,24
	Endogenous stratification	$\mu = \frac{H}{Q}\left[\frac{1-H}{1-Q} + \left(\frac{H}{Q} - \frac{1-H}{1-Q}\right) G(x\theta)\right]^{-1} G(x\theta)$, $(Q,H) = \{(0.9,0.7),(0.9,0.5),(0.5,0.3),(0.5,0.1)\}$	25,26,27,28

* $m(y,x^*) = \frac{0.5\theta_t^2}{G(x^*\theta)}\left[\nabla^2_{x\theta} G(x^*\theta) + \frac{2}{\theta_t}\nabla_{x\theta} G(x^*\theta)\, l_X(x^*)\right]$, where $\nabla_{x\theta}$ denotes derivative with respect to $x\theta$, $l_x(x^*) = \ln f(x^*)$, and θ_t is the coefficient associated to the variable measured with error.

Figure 1: RESET1 and RESET2 versus Sup-RESET