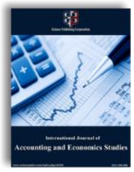
International Journal of Accounting and Economics Studies, 6 (2) (2018) 53-55

International Journal of Accounting and Economics Studies

Website: www.sciencepubco.com/index.php/IJAES
doi: 10.14419/ijaes.v6i2.13478
Research paper



A regression error specification test (RESET) for the truncated regression model

Sunil Sapra *

Department of Economics and Statistics California State University Los Angeles, CA 90032 USA
**Corresponding author E-mail: ssapra@calstatela.edu*

Abstract

While a variety of specification tests are routinely employed to test for misspecification in linear regression model, such tests and their applications to the truncated and censored regression models are uncommon. This paper develops a regression error specification test (RESET) for the truncated regression model as an extension of the popular RESET for the linear regression model (Ramsey (1969)). The two proposed extensions TRESET1 and TRESET2 developed in the paper are applied to labor force participation data from Mroz (1987). The paper studies the empirical size and power properties of the proposed tests via Monte Carlo experiments. Our simulation results suggest that both TRESET tests have reasonably good size and power properties for the truncated regression model in medium to large samples. However, TRESET2 consistently outperforms TRESET1 both in terms of empirical size and power in our experiments.

Keywords: *Truncated Regression; RESET; TRESET; Empirical Size; Empirical Power.*

1. Introduction

Misspecification can lead to serious bias and inefficiency in estimation of micro-econometric models. Ramsey (1969) developed a regression error specification test (RESET) to test for misspecification in regression models. RESET is a popular test, which is routinely employed to detect omitted variables and incorrect functional form in the linear regression model. It uses an artificial regression, which includes the predicted value of the dependent variable y and its higher powers among the regressors and tests the statistical significance of these terms.
In recent years, RESET-type tests have become popular in applied econometric research due to their computational convenience and good statistical properties. Previous studies by Godfrey and Orme (1994) and Horowitz (1994) reported good size and power properties for RESET tests for the Tobit model against functional form misspecification. Peters (2000) studied the performance of RESET tests for linear, Tobit, and Weibull regression models and emphasized that RESET-type tests are available for any econometric model for which the explanatory variables enter the model via a regression function. Sapra (2005) extended the RESET to generalized linear models (GLMs) in which the dependent variable is possibly continuous, categorical or count and explanatory variables enter the model via a link function. Nevertheless, applications of RESET-type specification tests to the truncated and censored regression models are uncommon despite their computational convenience and previous studies have not reported results on the performance of the RESET-type tests for the widely used truncated regression model. This paper extends the RESET approach to the truncated regression model and develops two RESET-type tests for the truncated regression model called TRESET1 and TRESET2. The TRESET tests are applied to Mroz (1987) data on labor force participation and their size and power properties are studied for the truncated regression model via simulation experiments.

This paper is organized as follows. Section 2 presents the two proposed TRESET tests for the truncated regression model. Section 3 presents an application of these tests to labor force participation data from Mroz (1987). Section 4 presents the results of Monte Carlo experiments on the size and power properties of the proposed TRESET tests. Section 5 provides concluding remarks.

2. TRESET for the truncated regression

Let y denote the dependent variable truncated from below at 0 and having the following relationship

$$y_i = \beta' x_i + \varepsilon_i, \quad (1)$$

β is a kx1vector of unknown parameters;
x_i is a kx1 vector of constants; ε_i are errors that are independently distributed as $N(0, \sigma^2)$. In the truncated regression model with truncation from below at 0, observations on both x and y are missing for $y_i \le 0$. The truncated mean is $f_i = E\big((y_i | y_i > 0)\big) = \beta' x_i - \frac{\sigma \phi_i}{\Phi_i}$ (Maddala (1983)), and $f_{2i} = f_i^2$ and $f_{3i} = f_i^3$. ϕ_i and Φ_i are respectively the density and the cumulative distribution functions of the standard normal random variable evaluated at $\beta' x_i / \sigma$. Let $\hat{f}_i$ denote the truncated mean f_i of y_i evaluated at the maximum likelihood estimates $\hat{\beta}$ and $\hat{\sigma}$ of β and σ respectively and $\widehat{f_{2i}} = \hat{f}_i^{\,2}$ and $\widehat{f_{3i}} = \hat{f}_i^{\,3}$.
Consider the following two artificial truncated regression models.

$$y_i = \beta' x_i + \gamma_1 \widehat{f_{2i}} + \varepsilon_i, \quad (2)$$

$$y_i = \beta' x_i + \gamma_1 f_{2i} + \gamma_2 \widehat{f_{3i}} + \varepsilon_i. \quad (3)$$

TRESET1

TRESET1 tests misspecification by testing the significance of $\widehat{f_{2i}}$ in equation (2) as follows.

H_0: $\gamma_1 = 0$ (No misspecification)

H_1: $\gamma_1 \neq 0$ (Misspecification due to omitted variables and/or incorrect functional form),

where γ_1 is the coefficient of $\widehat{f_{2i}}$ in equation (2).

TRESET2

TRESET2 tests misspecification by testing the joint significance of $\widehat{f_{2i}}$ and $\widehat{f_{3i}}$ in equation (3):

H_0: $\gamma_1 = \gamma_2 = 0$ (No misspecification)

H_1: $\gamma_1 \neq 0$ and/or $\gamma_2 \neq 0$ (Misspecification due to omitted variables and/or incorrect functional form) ,

where γ_1 is the coefficient of $\widehat{f_{2i}}$ and γ_2 is the coefficient of $\widehat{f_{3i}}$ in equation (3).

Each TRESET is a likelihood ratio test and TRESET1$\sim\chi^2(1)$ and TRESET2$\sim\chi^2(2)$ under H_0.

3. An empirical application

3.1. Application of TRESET1 and TRESET2 to Mroz data on labor supply

In this section, we present an empirical application of TRESET1 and TRESET2 tests. Our proposed TRESET tests compare a truncated regression model with no higher order terms with one with higher order terms. Specifically, TRESET1 compares a truncated regression model with no higher order terms with a truncated regression model with the second power of the predicted truncated response and TRESET2 compares the former with a truncated regression model with the third power of the predicted truncated response.

Mroz (1987) Data

The cross-sectional labor force participation data are from Mroz (1987) and are also available at the website for the text, Principles of Econometrics by Hill et al. (2013). The data consist of observations on 753 households. Truncation of HOURS from below at 0 led to a loss of 325 observations on dependent and independent variables. The dependent variable is HOURS and the independent variables are EDUC, EXPER.AGE, and KIDSLT6. Additional variables $\hat{f}_2$ and $\hat{f}_3$ are included in the truncated regressions for TRESET tests. Two TRESET tests are presented for the truncated regression model: TRESET1 and TRESET2. TRESET1 tests for misspecification by testing the null hypothesis of no misspecification against the alternative hypothesis of misspecification by testing that the coefficient of $\hat{f}_2$ is zero. Similarly, TRESET2 tests for misspecification by testing the null hypothesis of no misspecification against the alternative hypothesis of misspecification by testing that the coefficients of $\hat{f}_2$ and $\hat{f}_3$ are simultaneously equal to zero.

VARIABLE DEFINITIONS

HOURS = Wife's hours of work in 1975

KIDSLT6 = Number of children < 6 years old in household

AGE = Woman's age in years

EDUC = Wife's educational attainment in years

EXPER = Actual years of wife's previous labor market experience

The data characteristics are summarized in Table 1.

Table 1: Summary Statistics: Labor Supply Data from Mroz (1987), 753 Observations Source: 1976 Panel Study of Income Dynamics, Mroz (1987)

Variable	Observations	Mean	Std. Dev.	Min	Max
HOURS	753	740.5764	871.3142	0	4950
EDUC	753	12.28685	2.280246	5	17
EXPER	753	10.63081	8.06913	0	45
AGE	753	42.53785	8.072574	30	60
KIDSLT6	753	0.2377158	0.523959	0	3

Results of application of TRESET1 and TRESET2 to Mroz (1987) data on labor force participation

Tables 2 and 3 below present the results of application of TRESET1 and TRESET2 to Mroz (1987) data. At 5% significance level, TRESET1 rejects the null hypothesis H_0 and confirms misspecification while TRESET2 fails to reject the null hypothesis H_0 and indicates no misspecification. This suggests that higher order nonlinear terms beyond the quadratic functions of explanatory variables are not likely to improve functional form specification.

Table 2: Application of TRESET1 to Truncated Regression with Mroz Data Note: 325 Obs. Truncated.

Variable	Coefficient	Standard Error	z-statistic	p-value
Intercept	4232.143	1242.292	3.41	0.001
EDUC	-58.84169	28.66519	-2.05	0.040
EXPER	118.1376	37.19241	3.18	0.001
AGE	-58.45248	18.82623	-3.10	0.002
KIDSLT6	-1028.744	319.3163	-3.22	0.001
f_2	-0.000745	0.0003766	-1.98	0.048

TRESET1 = 2 (3394.9808 – 3392.9598) = 4.042, p-value (Prob > chi2 (1)) = 0.044381, significant at α = 0.05.

Table 3: Application of TRESET2 to Truncated Regression with Mroz Data Note: 325 Obs. Truncated.

Variable	Coefficient	Standard Error	z-statistic	p-value
Intercept	9804.668	4931.217	1.99	0.047
EDUC	-112.7685	54.62696	-2.06	0.039
EXPER	234.6875	107.0877	2.19	0.028
AGE	-122.2268	58.03448	-2.11	0.035
KIDSLT6	-2202.012	1059.947	-2.08	0.038
f_2	-0,004321	0.0030713	-1.41	0.159
f_3	0.0000011	0.000000936	1.18	0.239

TRESET2 = 2(3394.9808 – 3392.2602) = 5.4412, p-value (Prob > chi2(2)) = 0.065835, not significant at α =0.05.

4. Monte Carlo simulations

This section presents results of simulation experiments on the empirical size and power properties of TRESET1 and TRESET2 for the truncated regression model.

Computation of Empirical Size

TRESET1

TRESET1 tests misspecification by testing the significance of $\widehat{f_{2i}}$ in equation (2) as follows.

H_0: $\gamma_1 = 0$ (No misspecification)

H_1: $\gamma_1 \neq 0$,

where γ_1 is the coefficient of $\widehat{f_{2i}}$ in equation (2).

In order to estimate the empirical size of the TRESET1 test, 1000 samples of size 25, 100, 200, 500, and 1,000 on the variable y were generated under the null hypothesis H_0 according to the following truncated regression model

$$y^* = x^* - 6 + \varepsilon,$$

where $x^* \sim U(0,5)$ and $\varepsilon \sim N(0,16)$. The sample on the right-hand side variable x* was generated according to the Uniform law U(0,5) and held fixed once it was generated. The truncated sample (x, y) was then generated by retaining only those (x*, y*) for which y *> 0 and dropping the remaining cases.

The model under the null hypothesis was model (1) assuming no misspecification so that no higher degree terms in x are needed.

TRESET2

TRESET2 tests misspecification by testing the joint significance of $\widehat{f_{2i}}$ and $\widehat{f_{3i}}$ in equation (3):

H_0: $\gamma_1 = \gamma_2 = 0$

H_1: $\gamma_1 \neq 0$ and/or $\gamma_2 \neq 0$,

where γ_1 is the coefficient of $\widehat{f_{2i}}$ in equation (2) and γ_2 is the coefficient of $\widehat{f_{3i}}$ in equation (3).

Results on the empirical size properties of TRESET1 and TRESET2 are presented in Table 4. As the sample size increases, the empirical size of each of these tests approaches the nominal significance level of 0.05. The empirical size of TRESET1 decreases steadily from 0.34 to 0.084 as the sample size increases from 50 to 1,000 with the sole exception of n = 1,000, when it increases to 0.084 from 0.054. Similarly, the empirical size of TRESET2 decreases steadily from 0.27 to 0.066 as the sample size increases from 50 to 1,000 with the sole exception of n = 1,000, when it increases to 0.066 from 0.046. Nevertheless, for both TRESET1 and TRESET2, the empirical size stays close to the nominal significance level $\alpha = 0.05$ for samples of size 100 or greater. Furthermore, in terms of empirical size, TRESET2 outperforms TRESET1 in that its empirical size is smaller than that of TRESET1 for all sample sizes and is closer to the actual size of the test relative to the empirical size of TRESET1.

Table 4: Empirical Size Properties of TRESET1 and TRESET2 for $\alpha = 0.05$

Sample Size (n)	Empirical Size of TRESET1	Empirical Size of TRESET 2
50	0.34	0.27
100	0.076	0.064
200	0.068	0.058
500	0.054	0.046
1,000	0.084	0.066

Computation of Empirical Power of TRESET1 and TRESET2
TRESET1
H_0: $\gamma_1 = 0$
H_1: $\gamma_1 \neq 0$,
Where γ_1 is the coefficient of $\widehat{f_{2i}}$ in equation (2).
In order to compute empirical power for TRESET1, 1000 samples of sizes 50, 100, 200, 500, and 1,000 on the variable y were generated under the alternative hypothesis H_1 according to the following truncated regression model

$$y^* = x^* + 0.5x^{*^2} - 10 + \varepsilon,$$

where $x^* \sim U(0,5)$ and $\varepsilon \sim N(0,16)$. the sample on the right-hand side variable x* was held fixed once it was generated. The truncated sample (x, y) was then generated by retaining only those (x*, y*) for which y *> 0 and dropping the remaining cases. The model under the null hypothesis was model (1) assuming no misspecification so that no higher degree terms in x are needed.
TRESET2
TRESET2 tests misspecification by testing the joint significance of $\widehat{f_{21}}$ and $\widehat{f_{31}}$ in equation (3):
H_0: $\gamma_1 = \gamma_2 = 0$
H_1: $H_1: \gamma_1 \neq 0$ and/or $\gamma_2 \neq 0$,
Where γ_1 is the coefficient of $\widehat{f_{21}}$ in equation (2) and γ_2 is the coefficient of $\widehat{f_{31}}$ in equation (3).
In order to compute empirical power for TRESET2, 1000 samples of sizes 50, 100, 200, 500, and 1,000 on the variable y were generated under the alternative hypothesis H_1 according to the following truncated regression model

$$y^* = x^* + 0{,}2x^{*^2} + 0.4x^{*^3} - 12 + \varepsilon,$$

where $x^* \sim U(0,5)$ and $\varepsilon \sim N(0,16)$. the sample on the right-hand side variable x* was held fixed once it was generated. The truncated sample (x, y) was then generated by retaining only those (x*, y*) for which y *> 0 and dropping the remaining cases. The model under the null hypothesis was model (1) assuming no misspecification so that no higher degree terms in x are needed.
Results on the empirical power properties of TRESET1 and TRESET2 are presented in Table 5. The empirical power of TRESET1 increases steadily from 0.102 to 0.798 as the sample size increases from 50 to 1,000. The empirical power of TRESET2 increases steadily from 0.156 to 1.0 as the sample size increases from 50 to 1,000. Furthermore, the empirical power of TRESET2 is consistently higher than that of TRESET1 for all sample sizes.

Table 5: Empirical Power Properties of TRESET1 and TRESET2 for $\alpha = 0.05$

Sample Size (n)	Empirical Power of TRESET1	Empirical Power of TRESET 2
50	0.102	0.156
100	0.122	0.294
200	0.226	0.606
500	0.472	0.968
1,000	0.798	1.0

It is apparent that TRESET2 outperforms TRESET1 for the truncated regression model, both in terms of empirical size and power properties.

5. Conclusion

Motivated by the computational convenience and good statistical properties of the regression error specification tests (RESET) for misspecification in linear regression model, this paper has developed two versions of the RESET for the truncated regression model: TRESET1 and TRESET2. The tests were applied to Mroz (1987) data on labor force participation and their properties were studied via simulation experiments. Our results suggest that the empirical size for each test is very close to the nominated significance level. Furthermore, both tests have reasonable power properties in medium to large samples although TRESET2 outperforms TRESET1 both in terms of empirical size and empirical power in medium to large samples. These tests are computationally convenient and require only the predicted value of the dependent variable y under truncation and the maximum values of the log-likelihood functions under the null and alternative hypotheses, which can be easily computed using common econometric and statistical software packages. Routine use of TRESET1 and TRESET2 for detecting misspecification can lead to potential improvements in the fitted truncated regression models. If misspecification is confirmed, model performance may be improved by including higher degree polynomials of predictor variables or by using semi-parametric models, such as the generalized additive models (Hastie and Tibshirani (1990)).

References

[1] Godfrey LG & Orme, CD (1994) the sensitivity of some general checks to omitted variables in the linear model. International Economic Review 35, 489-506. https://doi.org/10.2307/2527066.
[2] Hastie T & Tibshirani R (1990) Generalized Additive Models, Chapman and Hall/CRC Monographs on Statistics and Probability, Boca Raton.
[3] Hill R C, Griffiths WA & Lim G (2013) Principles of Econometrics. Wiley, New York.
[4] Horowitz JL (1994) Bootstrap-based critical values for the Information Matrix Test, Journal of Econometrics 38, 375-386. https://doi.org/10.1016/0304-4076(94)90092-2.
[5] Horowitz JL & Neumann GR (1989) Specification Testing in censored regression models: parametric and semi-parametric methods, Journal of Applied Econometrics 4, S61-S86. https://doi.org/10.1002/jae.3950040505.
[6] Maddala GS (1983) Limited Dependent and qualitative variables in Econometrics. Cambridge University Press, New York. https://doi.org/10.1017/CBO9780511810176.
[7] Mroz T (1987) the sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. Econometrica 55, 765-799. https://doi.org/10.2307/1911029.
[8] Peters S (2000) on the use of the RESET tests in Microeconometric Models, Applied Economics Letters 7, 361-365. https://doi.org/10.1080/135048500351285.
[9] Ramsey JB (1969) Tests for specification errors in classical least-squares regression analysis, Journal of the Royal Statistical Society, Series B 31, 350-71.
[10] Sapra S (2005) A Regression Error Specification (RESET) Test for Generalized Linear Models, Economics Bulletin 3, AccessEcon, 1-6.