Rochester Center for
Economic Research

A New Regression Specification Error Test

Leung, Siu Fai and Shihti Yu

Working Paper No. 365
November 1993

University of
Rochester

A New Regression Specification Error Test

Siu Fai Leung and Shihti Yu

Rochester Center for Economic Research
Working Paper No. 365

A New Regression Specification Error Test

Siu Fai Leung

University of Rochester

and

Shihti Yu

National Chung Cheng University

November 1993

Correspondence To:
Prof. Siu Fai Leung
Department of Economics
University of Rochester
Rochester, NY 14627
U.S.A.

We thank Bruce Hansen, Adrian Pagan, Hashem Pesaran, and Charles Phelps for comments and suggestions, as well as Terence Chong for assistance. This research was supported by the NIAAA.

ABSTRACT: RESET and T-S RESET are two popular tests for specification error in the classical linear regression model. A major weakness of these two tests is that they have little statistical power if the test variables are not correlated with the omitted variables. In this paper we propose a new RESET that does not suffer from this deficiency. We create several dummy variables by categorizing the data into a few groups according to the values of the included regressors, and then use the dummy variables as the test variables. If the functional form is incorrect or some regressors are omitted, then the means of the omitted variables differ across the groups. An F-test on the coefficients of the dummy variables will detect the specification error if the difference is large enough. The strength of the dummy variable test (DVT) is that it does not rely on the correlations between the omitted variables and the included regressors. Our Monte Carlo study shows that, although DVT does not completely dominate RESET and T-S RESET, it offers a promising alternative to the two tests. When RESET or T-S RESET performs well, DVT performs equally well in most cases. On the other hand, when RESET and T-S RESET fail, DVT still has considerable power in detecting the specification error.

KEY WORDS: Dummy variable; Incorrect functional form; Monte Carlo Experiment; Omitted variable; RESET; Specification error test; T-S RESET.

Siu Fai Leung
Assistant Professor
Department of Economics
University of Rochester
Rochester, NY 14627
U.S.A.

Shihti Yu
Associate Professor
Institute of International Economics
National Chung Cheng University
Chia-Yi 621
TAIWAN

1. INTRODUCTION

In the classical linear regression model, incorrect functional form or omitted variables may result in biased estimates and invalid hypothesis testing. Even if the omitted variable is uncorrelated with every included variable, so that the coefficient estimates are unbiased and consistent, hypothesis testing remains invalid because the intercept estimate is still biased. Any forecast based on the estimates is also biased. Statistical tests for these types of specification errors have been developed, the most popular of which is Ramsey's (1969) RESET (regression specification error test).[1]

In addition to Ramsey's RESET, Thursby and Schmidt (1977) propose a variant of RESET, T-S RESET, which differs from RESET only in the set of test variables. While RESET uses the powers of the predicted value of the dependent variable as the test variables, T-S RESET employs the powers of the included regressors. Thursby and Schmidt's Monte Carlo study suggests that T-S RESET is superior to RESET and several other choices of test variables.

A major weakness of RESET and T-S RESET is that they have little statistical power if the test variables are not correlated with the omitted variables. In this paper we propose a new RESET that does not suffer from this deficiency. We create several dummy variables by categorizing the data into a few groups according to the values of the included regressors, and then use the dummy variables as the test variables. If the

---

[1] RESET is now available in some econometric software packages, e.g., MICROFIT 3.0 (Pesaran and Pesaran 1991). The test can be invoked by a single command, making it extremely easy to use.

functional form is incorrect or some regressors are omitted, then the means of the omitted variables differ across the groups. An F-test on the coefficients of the dummy variables will detect the specification error if the difference is large enough. Our test, which we label DVT for convenience, is a variant of RESET because the only difference is in the set of test variables. The strength of DVT is that it does not rely on the correlations between the omitted variables and the included regressors.

We carried out a series of Monte Carlo experiments to evaluate the three tests. We find that DVT performs well when RESET and T-S RESET fail. Even when RESET and T-S RESET perform well, DVT performs better or equally well in some cases, and it is only slightly inferior in others. The experiments show that, although DVT does not completely dominate RESET and T-S RESET, it serves as an excellent complement (if not substitute) for RESET and T-S RESET. A byproduct of our experiments is that, in contrast to Thursby and Schmidt's (1977) findings, T-S RESET is not always superior to RESET.

The plan of the paper is as follows. Section 2 gives a brief review of RESET and T-S RESET. Section 3 introduces DVT. Section 4 describes the designs of the experiments and Section 5 reports the simulation results. Section 6 concludes the paper.

2. Review of RESET and T-S RESET

Consider the standard classical linear regression model

$$y = X\beta + u, \quad (1)$$

where y is an Nx1 vector of observations on the dependent variable, $X = (x_1\ x_2\ \ldots\ x_K)$ is an NxK matrix of regressors, and u is an Nx1 vector of normally distributed error terms. Both RESET and T-S RESET seek to test the null hypothesis $E(u|X) = 0$ versus the alternative $E(u|X) \neq 0$. Hence, the tests are designed to check for specification errors such as omitted variables, incorrect functional form, or nonzero correlation between X and u. Ramsey (1969) assumes that, under the alternative, $E(u|X)$ can be approximated by a linear combination of some observable variables Z, i.e., $E(u|X) = Z\theta$ (Z is an NxG matrix and θ a Gx1 vector), so one can test for specification error by testing whether $\theta = 0$. Ramsey and Schmidt (1976) show that Ramsey's (1969) RESET test for these types of specification errors is equivalent to testing whether $\theta = 0$ in the augmented regression $y = X\beta + Z\theta + u$.

Since Z is unknown to the researcher, the main issue is what proxies should be used as the test variables. Let W denote a matrix of test variables, then the regression

$$y = X\beta + W\gamma + u \tag{2}$$

gives $\hat{\gamma} = (W'M_XW)^{-1}W'M_Xy$, where $M_X = I-X(X'X)^{-1}X'$. Under the null hypothesis, $E(\hat{\gamma}) = 0$. Under the alternative, $E(\hat{\gamma}) = (W'M_XW)^{-1}W'M_XZ\theta$. Therefore, RESET has power against the null hypothesis if either W is correlated with Z; or if W is uncorrelated with Z, both W and Z are correlated with X (Thursby and Schmidt 1977). If either one of these two conditions holds, then $W'M_XZ\theta \neq 0$, hence $E(\hat{\gamma}) \neq 0$.

Ramsey (1969) suggests using the powers of $\hat{y}$ as the test variables, i.e., $W = [\hat{y}^{(2)},\hat{y}^{(3)},\hat{y}^{(4)},\ldots]$, where $\hat{y} = X\hat{\beta}$ and $\hat{y}^{(j)}$ (j = 2,3,4,...) denotes the Nx1 vector obtained from raising each element of $\hat{y}$ to the jth power. In their Monte Carlo study, Ramsey and Gilbert (1972) find that $W = [\hat{y}^{(2)}, \hat{y}^{(3)}, \hat{y}^{(4)}]$ is sufficient for practical purposes. Thursby and Schmidt (1977) propose using the powers of the regressors as the test variables, i.e., $W = [X^{(2)}, X^{(3)}, X^{(4)}, \ldots]$, where $X^{(j)} = (x_1^{(j)}\ x_2^{(j)} \ldots x_K^{(j)})$ if x_1 is not a column of ones, and $X^{(j)} = (x_2^{(j)} \ldots x_K^{(j)})$ if x_1 is a column of ones. They conduct a Monte Carlo experiment to compare four different sets of test variables (including Ramsey's) and find that the powers of the regressors are the best.

3. DUMMY VARIABLE TEST

By making use of dummy variables, we propose an alternative to RESET and T-S RESET. Let c be a scalar variable with observed range $[\underline{c}, \bar{c}]$. For instance, c may be one of the regressors. Suppose c is categorized into Q intervals $[h_0,h_1)$, $[h_1,h_2)$, ..., $[h_{Q-1},h_Q)$, where $h_0 = \underline{c}$ and $h_Q = \bar{c}$. Define Q dummy variables d_{iq} ($i=1,2,\ldots,N$; $q=1,2,\ldots,Q$) as follows:

$$d_{iq} = 1 \quad \text{if } c_i \in [h_{q-1},h_q),$$
$$= 0 \quad \text{otherwise,}$$

and $d_{iQ} = 1$ if $c_i = h_Q$. Let $d_i = (d_{i,1},\ldots,d_{i,Q-1})$, $D = (d_1',\ldots,d_N')'$, and $\gamma = (\gamma_1,\ldots,\gamma_{Q-1})'$. We then consider the regression model

$$y = X\beta + D\gamma + u. \tag{3}$$

The last dummy, d_{iQ}, is excluded from D in order to avoid perfect collinearity in (3).

To run DVT, one must choose a categorizing variable and decide the number and the length of the intervals. On the choice of the test variables, Thursby and Schmidt (1977, p.637) have rightly remarked that "we have little choice except to make the matrix Z of test variables depend on X in one way or another." Given a limited choice, the categorizing variable c has to be chosen from the set $C = \{x_1, x_2, \ldots, x_K, \hat{y}\}$. We suggest two ways to categorize the data. First, for each $c \in C$, the intervals $[h_{q-1}, h_q)$ ($q = 1,2,\ldots,Q$) are evenly spaced, i.e., $h_q - h_{q-1} = h_{q+1} - h_q$, with $h_0 = \underline{c}$ and $h_Q = \bar{c}$ ($[\underline{c}, \bar{c}]$ is the observed range of c). For convenience, we call this Categorization Method I. In practice, we find that DVT has reasonable power for Q = 2, 3, 4. If there are too few observations in some of the intervals, DVT does not work well, so an alternative way of categorizing the data is needed. In these situations we rank the observations in ascending (or descending) order of c and then divide them into Q groups such that each group has the same number of observations (or roughly the same number of observations if N/Q is not an integer). We label this Categorization Method II.[2]

It is possible that DVT may have low power against the null for a particular Q. We therefore suggest using several values of Q. In practice, we increase Q from 2 to 4, and we find that the test has quite good power against the null for at least one of the three values of Q. If DVT rejects the null hypothesis at a particular Q, then it is unnecessary to continue the test for higher values of Q.

The advantage of DVT over RESET and T-S RESET is that DVT does not

[2] One reason for including $\hat{y}$ in the set of categorizing variables is that if all the regressors $\{x_1, x_2, \ldots, x_K\}$ are dummy variables, then the regressors cannot be used as categorizing variables. In this case, the only available categorizing variable is $\hat{y}$.

depend on the correlation between Z and X. To compare these three tests, let $W_1 = [\hat{y}^{(2)},\hat{y}^{(3)},\hat{y}^{(4)}]$, $W_2 = [X^{(2)},X^{(3)},X^{(4)}]$, and $W_3 = D$. Assuming that the first column of X is a column of ones, then we can express more conveniently the dependent variable and the regressors as deviations from the means

$$E(\hat{\gamma}) = (W^{*\prime}M_{X^*}W^*)^{-1}W^{*\prime}M_{X^*}Z^*\theta, \tag{4}$$

where the asterisks denote that the elements of the matrix are deviations from the means. Consider the special case in which $W_1^{*\prime}Z^* = 0$, $W_2^{*\prime}Z^* = 0$, and $X^{*\prime}Z^* = 0$, then $E(\hat{\gamma}) = 0$ for RESET and T-S RESET. For DVT, $E(\hat{\gamma}) = (D^{*\prime}M_{X^*}D^*)^{-1}D^{*\prime}Z^*\theta$. Let a_q ($q = 1,2,\ldots,Q-1$) denote the qth element of the (Q-1)x1 vector $D^{*\prime}Z^*\theta$, then it can be shown that

$$a_q = \sum_{g=1}^{G} \theta_g \sum_{i=1}^{N} (d_{iq} - \bar{d}_q)(z_{ig} - \bar{z}_g),$$

where $\bar{d}_q$ and $\bar{z}_g$ are the means of d_{iq} and z_{ig} ($i=1,2,\ldots,N$), respectively. Since $\sum_{i=1}^{N} (d_{iq} - \bar{d}_q)(z_{ig} - \bar{z}_g) = \sum_{i=1}^{N} d_{iq}(z_{ig} - \bar{z}_g) = N^{q1}(\bar{z}_g^{q1} - \bar{z}_g) = N^{q1}N^{q0}(\bar{z}_g^{q1} - \bar{z}_g^{q0})/N$, where N^{qt} ($t = 0,1$) is the number of observations in which $d_{iq} = t$ and $\bar{z}_g^{qt}$ is the mean of z_{ig} of these observations, therefore

$$a_q = \sum_{g=1}^{G} \theta_g N^{q1}N^{q0}(\bar{z}_g^{q1} - \bar{z}_g^{q0})/N.$$

Even though $W_1^{*\prime}Z^* = 0$, $W_2^{*\prime}Z^* = 0$, and $X^{*\prime}Z^* = 0$, a_q is not necessarily zero. The power of DVT comes from the grouping of the observations. Each dummy variable separates the sample into two groups. If there are omitted variables, the means of the omitted variables may differ between the two groups. DVT will detect the specification error if the difference in the means of the omitted variables is large enough.

A simple example can be used to illustrate the point. Let K = 2 and G = 1, i.e., $y_i = \beta_1 + \beta_2 x_i + u_i$ (i=1,2,...,N). The true model is $y_i = \beta_1 + \beta_2 x_i + \theta z_i + u_i$. Assume that $x_i \sim U(-0.9^{1/2}, 0.9^{1/2})$,[3] and $z_i = x_i^5 - x_i^3 + 27x_i/140$. One can verify that $cov(z_i, x_i^j) = 0$, j = 1,2,3,4. Since z_i is uncorrelated with x_i^j (j = 1,2,3,4), $W_1^{*\prime}Z^*$, $W_2^{*\prime}Z^*$, and $X^{*\prime}Z^*$ are likely to be close to zero. As a result, both RESET and T-S RESET will have very low power.

For DVT, if Q = 2, then $a_1 = \theta N^{11}N^{10}(\bar{z}^{11} - \bar{z}^{10})/N$. Suppose x is the categorizing variable. As shown in Figure 1, Q = 2 means that $[h_0, h_1) = [-0.9^{1/2}, 0)$ and $[h_1, h_2] = [0, 0.9^{1/2}]$, i.e., $d_{i1} = 1$ if $x_i \in [-0.9^{1/2}, 0)$ and 0 otherwise; $d_{i2} = 1$ if $x_i \in [0, 0.9^{1/2}]$ and 0 otherwise. Consider the expression

$$\Delta = E(zI_{[-0.9^{1/2},0)}(x))/E(I_{[-0.9^{1/2},0)}(x)) - E(zI_{[0,0.9^{1/2}]}(x))/E(I_{[0,0.9^{1/2}]}(x)),$$

where $I_A(x)$ (an indicator function) equals 1 if $x \in A$, and 0 otherwise. One can verify that $E(zI_{[-0.9^{1/2},0)}(x)) = -E(zI_{[0,0.9^{1/2}]}(x)) = -\int_0^{0.9^{1/2}} (x^5 - x^3 + 27x/140)/[2(0.9)^{1/2}]dx = -0.00305$. Thus, $\Delta = -0.0122$. Since $\bar{z}_1^{11} - \bar{z}_1^{10}$ is the sample analog of Δ,[4] therefore $\bar{z}_1^{11} - \bar{z}_1^{10} \neq 0$. Hence, $a_1 \neq 0$ and DVT should have good power against the null when Q = 2. An important feature of this example is that DVT works well even though the omitted variable z is uncorrelated with the categorizing variable x at all.

DVT may fail to have power against the null for some particular values of Q. For

---

[3] Throughout the paper, $x_i \sim U(s_1, s_2)$ means that x_i has a uniform distribution with range $[s_1, s_2]$.

[4] It can be shown that, under standard regularity conditions, $\bar{z}_1^{11} - \bar{z}_1^{10}$ converges to Δ in probability.

instance, suppose that the true model is $y_i = \beta_1 + \beta_2 x_i + \theta z_i + u_i$ and the null hypothesis is $y_i = \beta_1 + \beta_2 x_i + u_i$, where $x_i \sim U(-1,1)$ and $z_i = x_i^2$. If Q = 2 and x is the categorizing variable, then the expression

$$\Delta^* = E(zI_{[-1,0)}(x))/E(I_{[-1,0)}(x)) - E(zI_{[0,1]}(x))/E(I_{[0,1]}(x))$$

is equal to zero because $E(x^2 I_{[-1,0)}(x)) = E(x^2 I_{[0,1]}(x))$. As a result, $\bar{z}_1^{11} - \bar{z}_1^{10}$, the sample analog of Δ^*, is close to zero. It follows that a_1 is close to zero and DVT is likely to have low power in detecting the omitted variable z when Q = 2. The symmetry of z on [-1,0) and [0,1] nullifies the ability of DVT to generate two different means for z on these two intervals. If Q = 3, then $a_1 = \theta N^{11}N^{10}(\bar{z}^{11} - \bar{z}^{10})/N$ and $a_2 = \theta N^{21}N^{20}(\bar{z}^{21} - \bar{z}^{20})/N$. The intervals become $[h_0,h_1) = [-1,-⅓)$, $[h_1,h_2) = [-⅓,⅓)$, and $[h_2,h_3] = [⅓,1]$. One can easily verify that

$$\Delta_1 = E(zI_{[-1,-\frac{1}{3})}(x))/E(I_{[-1,-\frac{1}{3})}(x)) - E(zI_{[-\frac{1}{3},1]}(x))/E(I_{[-\frac{1}{3},1]}(x)) \neq 0, \text{ and}$$

$$\Delta_2 = E(zI_{[-\frac{1}{3},\frac{1}{3})}(x))/E(I_{[-\frac{1}{3},\frac{1}{3})}(x)) - E(zI_{[-1,-\frac{1}{3})\cup[\frac{1}{3},1]}(x))/E(I_{[-1,-\frac{1}{3})\cup[\frac{1}{3},1]}(x)) \neq 0.$$

Hence, $\bar{z}^{11} - \bar{z}^{10}$ (the sample analog of Δ_1) and $\bar{z}^{21} - \bar{z}^{20}$ (the sample analog of Δ_2), should also be different from zero. Thus, DVT should have good power against the null when Q = 3. In this case, z is no longer symmetric on [-1,-⅓) and [-⅓,1). This example explains why we recommend that one should run DVT for several values of Q.

The only case in which DVT fails for all values of Q is when z and x are independent. For any intervals A and B on which x is defined, independence of z and x implies that $E(zI_A(x)) = E(z)E(I_A(x))$ and $E(zI_B(x)) = E(z)E(I_B(x))$. Consequently, $E(zI_A(x))/E(I_A(x)) - E(zI_B(x))/E(I_B(x)) = E(z) - E(z) = 0$, and DVT does not work well for any value of Q.[5] Of course, in this case both RESET and T-S RESET also have little power because z and x are uncorrelated.

Although the above theoretical investigation demonstrates that DVT is more powerful than RESET and T-S RESET in some special cases, its performance in more general cases remains to be examined. As theoretical comparisons of these three tests are intractable, we have to rely on Monte Carlo experiments to examine their relative effectiveness.

4. MONTE CARLO DESIGN

We conducted an extensive Monte Carlo investigation with a large number of models, and for brevity, we only report the results of 15 specifications here. Table 1 describes the models and the data generating processes. Two types of specification errors are studied: incorrect functional forms in Models 2 and 3, and omitted variables in Models 4 through 15. There is no specification error in Model 1. There are 1000 replications in each experiment and two sample sizes for each model (N = 50 and 200).

The specifications of Models 1 - 8 are identical to the ones studied by Thursby and Schmidt (1977).[6] The data for x_1 and x_2 were taken from appendix A in Ramsey and

[5] The reverse implication is also true. If $E(zI_A(x))/E(I_A(x)) - E(zI_B(x))/E(I_B(x)) = 0$ for any sub-intervals A and B, then $E(zI_A(x))/E(I_A(x)) = H$ must hold for any sub-interval A, where H is a constant. This implies that $E\{[E(z|x)-H]I_A(x)\} = 0$ for arbitrary A, so $E(z|x)$ must be equal to H, and therefore z and x must be independent.

[6] Thursby and Schmidt's (1977) Models 1, 2, and 8 were in turn drawn from Ramsey and Gilbert (1972).

Gilbert (1972) and were repeated five and twenty times to obtain samples of sizes 50 and 200. Regressors x_6 and x_7 were drawn independently from U(-1.5,1.5).[7] The specifications of Models 6 - 8 are identical except that the coefficients of the omitted variables are different. Models 9 - 15 are similar to Models 4 and 5; the only difference is the specification of the omitted variables. The omitted variables in Models 9 - 11 are cross products of the included regressors x_6 and x_7 (namely, x_6/x_7, x_6x_7, $x_6^2x_7$). In Models 12 and 13, the omitted variables are powers of x_6 (namely, $1/x_6, x_6^2$). Model 14 considers the case in which the omitted variable is a dummy variable whose value depends on x_6. Finally, Model 15 examines the example discussed in Section 4: the omitted variable x_{15} is uncorrelated with the first four powers of the included regressor x_6^*, where x_6^* and x_7^* were independently drawn from $U(-0.9^{1/2}, 0.9^{1/2})$. The error terms in Models 1 - 15 are the same and they were drawn from a standard normal distribution N(0,1).[8]

The test variables are $[\hat{y}^{(2)}, \hat{y}^{(3)}, \hat{y}^{(4)}]$ for RESET and $[X^{(2)}, X^{(3)}, X^{(4)}]$ for T-S RESET. As there are always two regressors in the null specification in each of the 15 models, for convenience we label the first and the second regressors generically as r_1 and

[7] The random variables x_6 and x_7 were generated using the GAUSS program (Aptech Systems 1992) with starting seeds 5598976 and 7900889. As Thursby and Schmidt (1977) did not detail how their x_5 and x_8 were generated (they just indicated that x_8 was a rearrangement of x_5, and x_6 and x_7 were orthogonal to x_5), we set $x_5 = x_6^2 + x_7^2$ and $x_8 = 6x_6 + x_7^2$. Similar to their design, our x_5 is orthogonal to x_6 and x_7. In their Models 4 and 5, the R^2s of the auxiliary regressions of $x_5 + u$ and $x_8 + u$ on x_6 and x_7 were 0 and 0.96, respectively. In our design, the corresponding R^2s were 0 and 0.97, closely matching their R^2s.

[8] The starting seeds for x_6^* and x_7^* are 5598976 and 7900889, respectively. The starting seeds for u_i are 10, 20, 30, ..., 10000 for the 1000 replications.

r_2. Accordingly, the set of categorizing variables for DVT is $\{r_1, r_2, \hat{y}\}$. We employed Categorization Method I to group the data. In some cases this method did not work well (serious collinearity problems due to insufficient observations in some of the intervals), so Categorization Method II was used. We set Q = 2, 3, 4 for r_1 and r_2, and Q = 2,..,5 for $\hat{y}$. All tests were performed at the five percent significance level. For misspecified models, the higher the rejection frequency, the better is the test. The rejection frequency indicates the power of the test. Since there is no misspecification in Model 1, the rejection frequency is expected to be around 5 percent.

5. SIMULATION RESULTS

The simulation results for sample sizes of 50 and 200 are given in Tables 2 and 3, respectively. The categorizing variables for DVT are r_1 and r_2 in both tables. Table 4 contains the results using $\hat{y}$ as the categorizing variable. The correlations between the omitted variables and the test variables are reported in Table 5.[9]

Model 1 has no misspecification and the rejection frequencies of RESET, T-S RESET, and DVT are all close to the size of the test (5 percent). All three tests have perfect power (100 percent rejection frequency) in detecting the incorrect functional forms in Models 2 and 3. For Models 4 and 5, DVT (Q = 4) outperforms RESET and performs almost as well as T-S RESET. All three tests have low power in Model 6; the rejection frequencies range from 5.0 percent to 6.2 percent when N = 50, and they are

[9] For brevity the correlations for Q = 4 and for $c = \hat{y}$ are not reported.

only slightly higher when N = 200. The tests fail to detect the omitted variable in Model 6 because the true coefficient of x_2 is 0.1, too small to be detectable. The power of the tests increases considerably in Model 7 when the coefficient of x_2 is raised to 1. The results in Model 7 show that DVT (Q = 4, c = r_2) and T-S RESET, which have similar rejection frequencies, are much more powerful than RESET when N = 50. RESET regains its power when N is increased to 200. In Model 8, RESET is still dominated by DVT and T-S RESET when N = 50. All three tests have 100 percent rejection frequency when N = 200.

The results from Models 1 - 8 verify Thursby and Schmidt's claim that T-S RESET performs better than RESET. DVT is also more powerful than RESET in Models 1 - 8. Although DVT does not dominate T-S RESET, its overall performance is as good as T-S RESET.

When the omitted variables are cross products of x_6 and x_7 (Models 9 - 11), T-S RESET is inferior to DVT. In Model 9, DVT dominates RESET and T-S RESET. Both RESET and T-S RESET completely fail to detect the omitted variable when N = 50. The rejection frequency of T-S RESET improves substantially when N is increased to 200. Although DVT is still more powerful than T-S RESET in Models 10 and 11, both tests are dominated by RESET. RESET performs well in these two models presumably because the test variables of RESET are composed of cross-products of x_6 and x_7. This finding invalidates Thursby and Schmidt's (1977) claim that the best set of test variables is the powers of the regressors.

Table 4 shows that DVT performs much better in Models 10 and 11 when the categorizing variable is $\hat{y}$. Although RESET is still superior to DVT in these two models, the rejection frequencies of DVT are significantly higher and are closer to those of RESET.

Perhaps the most striking result is the perfect power of DVT and the zero power of RESET and T-S RESET in detecting the omitted variable in Model 12. While DVT ($Q = 2$, $c = r_1$) achieves 100 percent rejection frequency for both sample sizes, RESET and T-S RESET have no power at all at the 5 percent significance level.[10] Table 5 shows that their failure can be attributed to the lack of correlations between the omitted variables and the test variables.

For Model 13, T-S RESET is the ideal test because the omitted variable x_{13} is the square of the included regressor x_6^2. It is therefore not surprising to find that T-S RESET attains 100 percent rejection frequency for both sample sizes. However, DVT ($Q = 3$ and $c = r_1$) also has the same perfect power as T-S RESET.[11] Even though x_6 and x_{13} are uncorrelated and the data are categorized according to x_6, DVT still accomplishes 100 percent rejection frequency. This confirms that DVT still has power even when there is no correlation between the omitted variable and the categorizing variable. Notice that DVT works very well when $Q = 3$ but not when $Q = 2$. In fact, DVT has very little

---

[10] To check the robustness of our results, we tried various modifications of x_{12} (e.g., $x_{12} = 10/x_6$, $30/x_6$, $50/x_6$, $1/x_6^3$, $1/x_6^5$). In all these cases DVT continues to substantially outperform RESET and T-S RESET.

[11] In addition to x_6^2, we also tried the cases $x_{13} = x_6^3$, x_6^4, and the results were similar.

power for both sample sizes when Q = 2. This corroborates the example in Section 4 that DVT does not work well when $z = x^2$ and Q = 2 because the omitted variable x_{13} is the square of the categorizing variable x_6.[12]

For Model 14, DVT is the ideal test because the omitted variable is a dummy variable whose value depends on x_6. The results confirm that DVT (Q = 2, $c = r_1$) dominates both RESET and T-S RESET for both sample sizes. While RESET remains to have low power, the rejection frequency of T-S RESET increases considerably (from 18.7 percent to 79.6 percent) when N increases from 50 to 200.[13]

Since the omitted variable x_{15} in Model 15 is uncorrelated with x_6, x_6^2, x_6^3, and x_6^4, RESET and T-S RESET are expected to have low power. This is confirmed in Tables 2 and 3, which show that DVT is much more powerful than RESET and T-S RESET. Although theoretically this design is the worst situation for T-S RESET, the simulation results reveal that the test still has some power. Unlike Model 12 in which T-S RESET has no power at all, the rejection frequencies are not zero (3.9 percent when N = 50 and 3.3 percent when N = 200). For Q = 2 and $c = r_1$, DVT has a remarkable increase in

---

12 When Q = 2, the intervals of x_6 are [-1.5,0) and [0,1.5], so x_{13} is symmetric on these two intervals. When Q = 3, the intervals become [-1.5,-0.5), [-0.5,0.5), and [0.5,1.5], so x_{13} is no longer symmetric on [-1.5,-0.5) and [-0.5,1.5].

13 We also studied a variation of Model 19 in which $x_{14} = 0$ if $x_6^2 \leq 0.75$, and $x_{14} = 1$ if $x_6^2 > 0.75$. In this case the categorizing variables of DVT (x_6 and x_7) are uncorrelated with the variable (x_6^2) that actually categorizes the data. We found that DVT is still more powerful than T-S RESET and RESET. The percentage rejection frequencies of DVT (Q = 4, $c = r_1$), T-S RESET, and RESET are 60.9, 39.2, and 9.7 when N = 50; and 98.4, 97.7, and 21.1 when N = 200.

power (from 1.4 to 95.6 percent) when the sample size is expanded to 200.[14]

Some general remarks are in order. First, Table 4 indicates that DVT performs quite well (especially in Models 10 and 11) when $\hat{y}$ is used as the categorizing variable. Second, the power of the tests can be quite sensitive to the sample size. For example, the rejection frequency of T-S RESET in Model 9 jumps from 0 to 87.8 percent when N increases from 50 to 200. Likewise, DVT ($Q = 2$ and $c = r_1$) shows substantial improvement in Model 15 when N increases to 200. RESET seems to be relatively more stable. Third, Table 5 shows that high correlations between the test variables and the omitted variables are not sufficient for RESET and T-S RESET to work. As shown in Models 6 - 8 in Tables 2 and 3, the power of the tests depend on other factors such as the coefficient of the omitted variable. On the other hand, holding other factors constant (such as the sample size and the magnitude of the coefficient of the omitted variable), if the correlations between the test variables and the omitted variables are uniformly low, then RESET and T-S RESET have very low power. For example, Table 5 shows that all of the three correlation coefficients for RESET are very small in Models 9, 12, 14, and 15, and the corresponding rejection frequencies in Table 3 are very low. Similarly, T-S RESET has very low power in Models 10, 12, and 15 because the correlation coefficients are also uniformly small.

[14] Unlike Models 9 - 14, the coefficient of the omitted variable x_{15} is 40 (instead of 1.5). A larger coefficient is employed in this model because the regressors x_6^* and x_7^* have smaller ranges than those of x_6 and x_7 in the previous models. We tried many different values (from 1.5 to 100) for the coefficient of x_{15} and we found that DVT always dominated RESET and T-S RESET.

In sum, our Monte Carlo study shows that, although DVT does not completely dominate RESET and T-S RESET, it offers a promising alternative to the two tests. When RESET or T-S RESET performs well, DVT performs equally well in most cases. On the other hand, when RESET and T-S RESET fail, DVT still has considerable power in detecting the specification error.

6. CONCLUSION

In this paper we propose a new regression test for specification error. Our dummy variable test builds on Ramsey's (1969) RESET and employs dummy variables as the test variables. The main advantage of DVT over RESET and T-S RESET is that it does not rely on the correlations between the omitted variables and the regressors. Although DVT may not be a perfect substitute for RESET and T-S RESET, our simulation results indicate that it is an excellent complement to the two tests.

The attractiveness of RESET lies in its ease of computation. Although DVT is computationally more involved than RESET and T-S RESET (especially when there are many explanatory variables), the implementation is not really time-consuming. Our experience is that DVT can easily be programmed into a single command in most statistical packages.

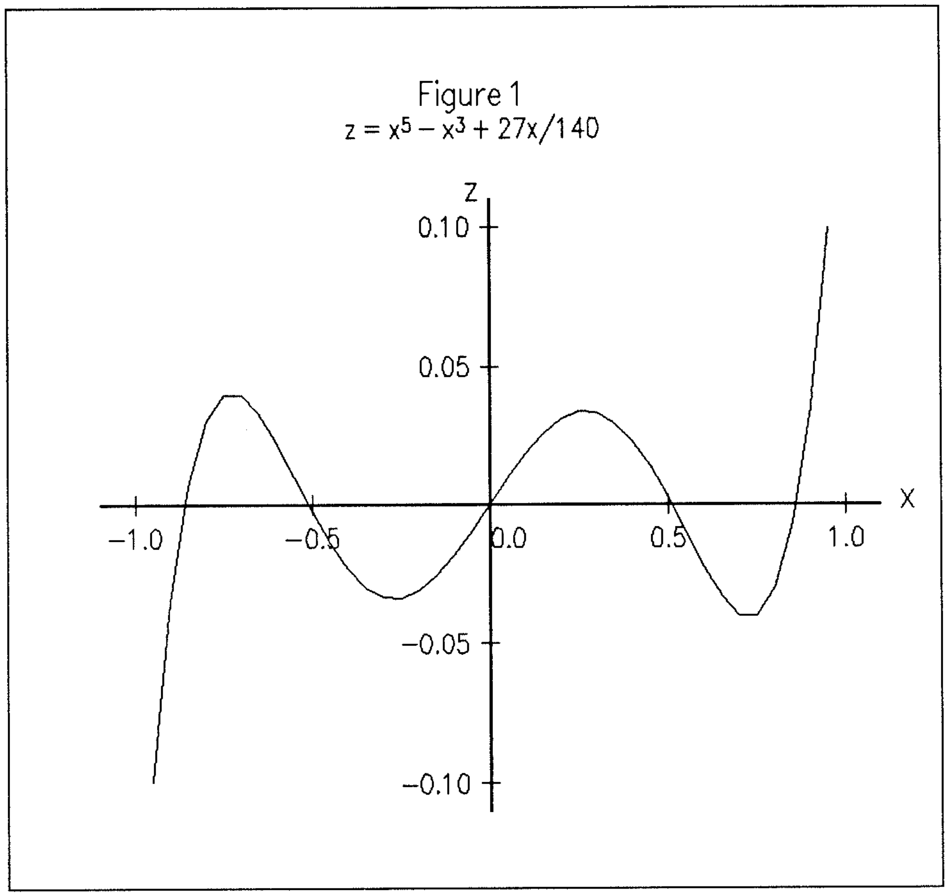
Figure 1
z = x5 – x3 + 27x/140
z
0.10
0.05
–0.05
–0.10
x
–1.0
–0.5
0.0
0.5
1.0

Table 1
Specifications of Models 1 - 15

Model	Specification		Regressors
	True	Null	
1	$y_i = 10 + 5\, x_{1i} - 2\, x_{2i} + u_i$	$y_i = \beta_0 + \beta_1\, x_{1i} + \beta_2\, x_{2i} + v_i$	$x_1 = (1.0, 3.7, 0.8, 9.9, 1.2, 6.6, 3.1, 8.5, 6.3, 7.3)$
2	$y_i = 1 + 2\, x_{1i} - 0.8\, x_{2i} + u_i$	$y_i = \beta_0 + \beta_1 \exp(x_{1i}/3) + \beta_2 \exp(x_{2i}/5) + v_i$	$x_2 = (9.0, 9.0, 8.0, 1.0, 8.0, 6.0, 6.0, 8.0, 12.0, 16.0)$
3	$y_i = x_{1i}^2 x_{2t}^{-0.8} + u_i$	$y_i = \beta_0 + \beta_1\, x_{1i} + \beta_2\, x_{2i} + v_i$	$x_3 = x_1 + x_2$
4	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{5i} + u_i$	$y_i = \beta_0 + \beta_1\, x_{6i} + \beta_2\, x_{7i} + v_i$	$x_4 = x_1^2/10$
5	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{8i} + u_i$		$x_5 = x_6^2 + x_7^2$
6	$y_i = 1 - 0.4\, x_{3i} + x_{4i} + 0.1\, x_{2i} + u_i$	$y_i = \beta_0 + \beta_1\, x_{3i} + \beta_2\, x_{4i} + v_i$	$x_6 \sim U(-1.5,1.5)$, $x_7 \sim U(-1.5, 1.5)$
7	$y_i = 1 - 0.4\, x_{3i} + x_{4i} + x_{2i} + u_i$		$x_8 = 6x_6 + x_7^2$
8	$y_i = 1 - 0.4\, x_{3i} + x_{4i} + 2\, x_{2i} + u_i$		$x_9 = x_6/x_7$
9	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{9i} + u_i$	$y_i = \beta_0 + \beta_1\, x_{6i} + \beta_2\, x_{7i} + v_i$	$x_{10} = x_6 x_7$
10	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{10i} + u_i$		$x_{11} = x_6^2 x_7$
11	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{11i} + u_i$		$x_{12} = 1/x_6$
12	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{12i} + u_i$		$x_{13} = x_6^2$
13	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{13i} + u_i$		$x_{14} = 1$ if $x_6 > 0$, $x_{14} = 1$ if $x_6 \leq 0$
14	$y_i = 0.8 - 0.6\, x_{6i} + x_{7i} + 1.5\, x_{14i} + u_i$		$x_{15} = x_6^{*5} - x_6^{*3} + 27x_6^*/140$
15	$y_i = 0.8 - 0.6\, x_{6i}^* + x_{7i}^* + 40\, x_{15i} + u_i$	$y_i = \beta_0 + \beta_1\, x_{6i}^* + \beta_2\, x_{7i}^* + v_i$	$x_6^* \sim U(-0.9^{½},0.9^{½})$, $x_7^* \sim U(-0.9^{½},0.9^{½})$

Table 2
Rejection Frequencies
Sample Size = 50
1000 replications

Model	RESET	T-S RESET	DVT					
			$Q = 2$ $c = r_1$	$Q = 2$ $c = r_2$	$Q = 3$ $c = r_1$	$Q = 3$ $c = r_2$	$Q = 4$ $c = r_1$	$Q = 4$ $c = r_2$
1	5.5	5.0	5.1	5.3	5.3	4.9	5.9*	5.1*
2	100.0	100.0	100.0	55.6	100.0	19.8	100.0*	100.0*
3	100.0	100.0	0.0	100.0	100.0	100.0	37.9*	100.0*
4	60.1	100.0	0.0	4.0	94.3	52.8	99.9	95.3
5	3.5	99.7	1.2	19.3	0.6	88.1	0.9	97.0
6	6.2	5.4	5.0	5.8	5.1	5.4	6.1*	5.9*
7	41.3	82.6	28.3	74.8	37.0	43.8	47.0*	83.6*
8	42.6	100.0	69.7	99.9	85.7	92.8	94.9*	100.0*
9	0.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0
10	92.7	2.8	1.9	1.4	16.5	0.0	3.9	0.1
11	5.1	0.2	1.2	1.3	1.4	0.2	0.9	2.6
12	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
13	21.0	100.0	0.2	0.4	100.0	0.1	100.0	1.9
14	4.1	16.4	66.5	7.1	5.9	4.4	53.8	6.6
15	1.0	3.9	1.4	1.4	87.1	2.7	99.6	0.3

Note: An asterisk indicates that Categorization Method II was used because of insufficient observations in some of the intervals.

Table 3
Rejection Frequencies
Sample Size = 200
1000 replications

Model	RESET	T-S RESET	DVT					
			$Q = 2$, $c = r_1$	$Q = 2$, $c = r_2$	$Q = 3$, $c = r_1$	$Q = 3$, $c = r_2$	$Q = 4$, $c = r_1$	$Q = 4$, $c = r_2$
1	4.0	5.0	4.4	4.1	4.4	4.3	3.8*	4.4*
2	100.0	100.0	100.0	100.0	100.0	100.0	100.0*	100.0*
3	100.0	100.0	0.0	100.0	100.0	100.0	100.0*	100.0*
4	100.0	100.0	0.5	4.6	100.0	100.0	100.0	100.0
5	2.8	100.0	0.3	5.8	2.8	100.0	0.2	100.0
6	7.8	7.2	6.2	8.2	5.9	5.9	6.4*	6.2*
7	98.6	100.0	90.9	100.0	98.1	99.1	99.6*	100.0*
8	100.0	100.0	100.0	100.0	100.0	100.0	100.0*	100.0*
9	0.0	87.8	0.0	100.0	0.0	98.2	0.0	100.0
10	100.0	1.2	0.2	0.4	5.8	0.0	0.3	0.0
11	29.7	1.6	6.1	0.6	0.7	1.4	6.7	1.9
12	0.0	0.0	100.0	42.9	0.0	0.0	100.0	0.0
13	91.6	100.0	2.7	1.7	100.0	12.6	100.0	4.0
14	3.3	76.9	100.0	4.0	5.5	3.8	99.8	3.2
15	3.3	3.3	95.6	19.6	91.8	2.1	100.0	4.1

Note: An asterisk indicates that Categorization Method II was used because of insufficient observations in some of the intervals.

Table 4
Rejection Frequency of DVT
Using $\hat{y}$ as the Categorizing Variable
1000 replications

Model	Sample Size = 50				Sample Size = 200			
	Q = 2	Q = 3	Q = 4	Q = 5	Q = 2	Q = 3	Q = 4	Q = 5
1	6.8	6.8	6.5	5.0	6.5	5.1	4.4	4.6
2	100.0	100.0	100.0	100.0*	100.0	100.0	100.0	100.0*
3	0.0	100.0	100.0	100.0*	66.5	100.0	100.0	100.0*
4	11.6	5.2	16.4	16.5	1.8	97.3	100.0	100.0
5	1.5	0.3	1.0	9.4	0.3	4.4	1.0	0.6
6	5.2	7.0	5.7*	5.5*	5.4	7.9	6.0*	5.9*
7	28.3	36.5	56.1	40.7*	90.9	98.0	99.8	96.0*
8	19.2	6.8	96.5*	88.1*	91.0	80.1	100.0	100.0*
9	0.0	0.0	61.3	82.5	0.0	1.9	0.3	0.0
10	2.2	38.4	61.6	74.3	1.1	99.5	100.0	99.9
11	11.2	28.9	6.6	11.2	6.3	12.6	13.5	9.4
12	2.4	0.0	2.1	2.2	48.4	0.0	0.0	0.0
13	1.7	7.1	1.9	3.3	1.8	1.4	41.6	69.4
14	4.4	4.6	3.8	4.3	4.5	3.6	4.0	5.8
15	1.0	7.0	0.8	6.5	6.2	3.9	4.8	4.7

Note: An asterisk indicates that Categorization Method II was used because of insufficient observations in some of the intervals.

Table 5
Correlations Between Test Variables and Omitted Variables
Sample Size = 200

Model	RESET			T-S RESET						DVT					
										Q = 2, $c = r_1$	Q = 2, $c = r_2$	Q = 3, $c = r_1$		Q = 3, $c = r_2$	
	$\hat{y}^2$	$\hat{y}^3$	$\hat{y}^4$	r_1^2	r_1^3	r_1^4	r_2^2	r_2^3	r_2^4	d_1	d_1	d_1	d_2	d_1	d_2
2	0.67	0.61	0.58	0.77	0.67	0.61	0.25	0.27	0.26	-0.32	-0.59	-0.46	0.48	-0.62	0.48
	-0.65	-0.66	-0.66	-0.54	-0.61	-0.64	0.78	0.73	0.70	0.46	-0.03	0.01	0.50	-0.24	0.33
3	0.98	0.99	0.99	0.77	0.85	0.90	-0.39	-0.25	-0.19	-0.36	-0.30	-0.36	-0.10	-0.13	-0.37
4	0.21	0.28	0.33	0.77	0.00	0.75	0.71	0.10	0.65	0.05	-0.12	0.27	-0.53	0.16	-0.53
5	0.49	0.85	0.54	0.00	0.92	0.02	0.08	-0.02	0.10	-0.83	0.01	-0.80	0.0	0.03	0.01
6	-0.65	-0.66	-0.65	0.76	0.78	0.78	-0.44	-0.55	-0.60	-0.59	0.01	0.59	0.23	-0.07	0.53
7	0.68	0.72	0.74	0.76	0.78	0.78	-0.44	-0.55	-0.60	-0.59	0.01	0.59	0.23	-0.07	0.53
8	0.89	0.86	0.83	0.76	0.78	0.78	-0.44	-0.55	-0.60	-0.59	0.01	0.59	0.23	-0.07	0.53
9	-0.09	0.06	-0.05	-0.02	-0.10	-0.02	0.15	0.00	0.10	-0.01	-0.14	-0.01	0.05	0.09	-0.19
10	-0.44	-0.43	-0.50	0.02	0.08	0.05	-0.04	-0.01	-0.07	-0.10	0.01	-0.13	0.01	0.0	0.04
11	0.47	0.61	0.45	0.04	0.03	0.01	0.04	0.68	0.03	0.03	-0.64	0.07	-0.07	-0.61	-0.05
12	0.08	0.08	0.05	0.00	0.05	0.00	0.05	0.04	0.03	-0.35	-0.13	-0.07	0.01	-0.02	-0.09
13	0.11	0.13	0.16	1.00	0.02	0.96	0.10	0.00	0.06	0.04	-0.06	0.36	-0.71	0.01	-0.14
14	0.08	0.09	0.09	-0.06	0.66	-0.04	-0.03	-0.02	-0.03	-0.99	0.02	-0.72	0.05	0.01	0.05
15	0.05	0.06	0.06	0.06	0.15	0.08	0.05	0.08	0.03	-0.26	0.02	0.01	-0.01	0.0	-0.02

REFERENCES

Aptech Systems (1992), GAUSS, Maple Valley, Washington: Aptech Systems, Inc.

Pesaran, M. Hashem and Pesaran, Bahram (1991), MICROFIT 3.0: An Interactive Econometric Software Package, Oxford: Oxford University Press.

Ramsey, James B. (1969), "Test for Specification Errors in Classical Linear Least Squares Regression Analysis," Journal of the Royal Statistical Society, Series B, 31, No. 2, 350-371.

Ramsey, James B. and Gilbert, Roy F. (1972), "A Monte Carlo Study of Some Small Sample Properties of Tests for Specification Error," Journal of the American Statistical Association, 67, 180-186.

Ramsey, James B. and Schmidt, Peter (1976), "Some Further Results on the Use of OLS and BLUE Residuals in Specification Error Tests," Journal of the American Statistical Association, 71, 389-390.

Thursby, Jerry G. and Schmidt, Peter (1977), "Some Properties of Tests for Specification Error in a Linear Regression Model," Journal of the American Statistical Association, 72, 635-641.