Research Report
Department of Statistics
Göteborg University
Sweden

The effect of non-normal error terms on the properties of systemwise RESET test

Ghazi Shukur

Research Report 1999:6
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Dept of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.handels.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

The Effect of Non-normal Error Terms on the Properties of Systemwise RESET Test

Ghazi Shukur
Department of Statistic
Göteborg University

ABSTRACT

The small sample properties of the systemwise RESET test for functional misspecification is investigated using normal and non-normal error terms. When using normally distributed or less heavy tailed error terms, we find the Rao's multivariate F-test to be best among all other alternative test methods. Using the bootstrap critical values, however, all test methods perform satisfactorally in almost all situations. However, the test methods perform extremely badly (even the *RAO* test) when the error terms are very heavy tailed.

Keywords: Systemwise Test of Functional Misspecification, Non-normal Error Terms, Small Sample Properties.

JEL Classification: C 32

1 Introduction

The RESET test proposed by Ramsey (1969) is a general misspecification test, which is designed to detect both omitted variables and inappropriate functional form. The RESET test is based on the Lagrange Multiplier principle and usually performed using the critical values of the F-distribution for single equation.

Shukur and Edgerton (1997), in what follows referred to as SE, have studied the properties of systemwise generalisations of Ramsey's RESET test for misspecification errors. They applied the systemwise Wald, Lagrange Multiplier (LM) and Likelihood Ratio (LR) tests to auxiliary regression systems. Various degrees-of-freedom corrections have been investigated, in particular the commonly used simple replacement of the number of observations by the degrees-of-freedom in the auxiliary regression and, for the LR test, the Edgeworth correction developed by Anderson (1958). The authors have also studied the properties of the systemwise F-test approximation proposed by Rao (1973). All in all eight different tests have been investigated.

The investigation has been carried out using Monte Carlo simulations. A large number of models are investigated, where the number of equations, degrees of freedom, error variance and stochastic properties of the exogenous variables have been varied. For each model SE have performed 10,000 replications and studied four different nominal sizes. The power properties have been investigated using 2,000 replications per model, where in addition to the properties mentioned above the degree of misspecification (measured as the relative differences in the explanatory power between the null and true models) and the correlation between the omitted and included variables have also varied. The SE's main conclusion is that the Rao's F-test is the best. The differences between the various versions of corrected LR tests are minimal when the number of equations (n) is small. The uncorrected LR test, and both the corrected and uncorrected Wald and *LM* tests, are shown to perform *extremely* badly in all situations. Only the *RAO* test performs satisfactorily when $n > 5$.

In view of these facts, Shukur and Mantalos (1997), in what follows referred to as SM, intended to improve the critical values of the test statistics, by employing the bootstrap technique, so that the size of the test approaches its nominal value. Given the bootstrap critical values they analysed the size and power of different generalization of the RESET test and compared the results with those found by SE. The SM analysis revealed that, in single equations, all test methods shown to be identical regarding size and power, while in system with many equations the eight tests reduced to three groups, namely Wald, LR (or Rao) and LM. Moreover, by using the bootstrap critical values, the analysis has also showed that, in almost all cases, the performances of all the tests are fairly satisfactory.

In both papers, however, the error terms have been generated from the normal distribution, but using a fat-tailed distribution could, of course, affect the properties of the tests. Evidence of fat-tailed distributions can often be found in empirical econometrics and/ or time series, e.g. finance, demand analysis, price expectations. If a test performs well under these conditions (i.e., when the underlying assumption that the error terms are normally and identically distributed may not hold), then it is usually referred to as robust. It is, hence, important to study the effect of fat tailed and/or the combined effect of fat tailed and functional misspecification on the properties of the RESET test.

The purpose of this paper is to investigate the properties of the systemwise RESET test in situation when the error terms are generated from non-normal distributions. In other words, the main point of this study is to investigate the ability of the RESET test to detect omitted variables when the error terms used in the models follow a non-normal distribution. More specifically, the error terms in this study will be generated by the t-distribution with degrees of freedom equal to: 1 (i.e. the Cauchy distribution), 2, 5, 10, 15 and 25. Proceeding in this manner, we cover a wide range and varying degree of the fatness in the tails of the errors.

In this study we use almost the same model specification and Monte Carlo design as in SE. The same eight tests, used in SE, are studied by Monte Carlo simulations. The properties of these tests have been investigated, once by using the critical values of the χ^2 and the F distributions, and once by using the bootstrap critical values. The results are then compared with those results from the two papers of SE and SM (i.e., when they use the normal distribution to generate the error terms and when using the bootstrap technique).

The paper is arranged as follows. In the next section we present the model which we analyse, and give the formal definition of a number of variants of the RESET test. In Section three we introduce the factors that can affect the small sample properties of the test, while the fourth Section presents the design of our Monte Carlo experiment. In Section five we describe the results concerning the size of the various tests, while power is studied in Section six. The conclusions of the paper are presented in the final section.

2 Model Specification

The regression model we will investigate is the same as in SE and SM. The model consists of n linear stochastic equations given by

$$\mathbf{Y}_t = \mathbf{X}_t \mathrm{B} + \varepsilon_t, \tag{1}$$

where Y_t and ε_t are $(1 \times n)$ vectors of endogenous variables and disturbances, X_t is a $(1 \times m)$ matrix of exogenous variables, B is a $(m \times n)$ matrix of parameters, and $t = 1,\ldots,T$. The data matrices Y and X are $(T \times n)$ and $(T \times m)$ respectively. The null hypothesis of correct specification implies that the error term will be independently and identically distributed conditional on the exogenous variables, that is

$$\varepsilon_t \mid \mathbf{X}_t \sim N(0, \Sigma_\varepsilon). \tag{2}$$

The hypothesis of correct functional form is equivalent to assuming that the disturbances have zero conditional mean, that is $H_0 : E(\varepsilon_t \mid \mathbf{X}_t) = 0$. The class of alternative hypotheses to this null is very general; omitted variables and incorrect functional form will obviously be members of the class.

The alternative hypothesis is specified through the following model

$$\mathbf{Y}_t = \mathbf{X}_t \mathrm{B} + \mathbf{Z}_t \Gamma + \varepsilon_t. \tag{3}$$

Z is in general unknown, and the tests that we will investigate estimate the following regression instead,

$$\mathbf{Y}_t = \mathbf{X}_t \mathrm{B} + \hat{\mathbf{Z}}_t \Gamma^* + \delta_t. \tag{4}$$

If the null hypothesis is correct, then $\Gamma = \Gamma^* = 0$ whatever the choice of $\hat{\mathbf{Z}}$. If the hypothesis is incorrect, then the choice of $\hat{\mathbf{Z}}$ will obviously affect the power of any test based on (4). The greater the correlation between $\hat{\mathbf{Z}}$ and the nonlinear part of the true conditional mean of Y, then, in general, the greater the power will be.

Ramsey (1969) proposed approximating the unknown conditional expectation of Y using a Taylor expansion around the conditional expectation *under the null hypothesis*, that is $X\beta$ (Ramsey was using a single equation, and β was thus a vector). Since the parameters are unknown, this was in turn approximated using $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathrm{b}}$, where $\hat{\mathrm{b}}$ was the OLS parameter estimate from the single equation version of (1). This is the RESET test procedure.

Let us now define a systemwise version of the RESET test. Following common terminology of double regression tests we will denote equation (1) as the *primary regression.* The first stage of the RESET test is performed by calculating the least squares' predictions from the primary regression, $\hat{\mathbf{Y}} = (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}$. These predictions are then used in the following *auxiliary regression,*[1]

$$\mathbf{Y}_t = \mathbf{X}_t\mathrm{B} + \hat{\mathbf{Y}}_t^2\Gamma_1^* + \hat{\mathbf{Y}}_t^3\Gamma_2^* \ldots + \hat{\mathbf{Y}}_t^{G+1}\Gamma_G^* + \delta_t \text{,} \tag{5}$$

where the *(t, i):th* elements of the power matrices are given by $[\hat{\mathbf{Y}}^j]_{ti} = \hat{y}_{ti}^j$. The RESET test is now performed by testing the hypothesis $H_0' : \Gamma_1^* = \cdots = \Gamma_G^* = 0$.

The practical implementation of the RESET test now depends on two factors. Firstly we must decide how many power matrices to include in the auxiliary regression (*i.e.*, determine *G*). Secondly we must decide which test method to use. In this paper, as in SE and SM, we will concentrate on the second question, and will let $G = 1$ throughout.[2]

Now, we denote by $\hat{\delta}_U$ the $(T \times n)$ matrix of estimated residuals from the *unrestricted* regression (5), and by $\hat{\delta}_R$ the equivalent matrix of residuals from the *restricted* regression with H_0 imposed. The matrix of cross-products of these residuals will also be defined as $S_U = \hat{\delta}_U'\hat{\delta}_U$ and $S_R = \hat{\delta}_R'\hat{\delta}_R$, and the Wald, Likelihood Ratio and Lagrange Multiplier test statistics are given by

$$W = T(\mathrm{tr}\mathbf{S}_U^{-1}\mathbf{S}_R - n) \text{,} \tag{6}$$

$$LR = T\ln U \text{, and} \tag{7}$$

$$LM = T(n - \mathrm{tr}\mathbf{S}_R^{-1}\mathbf{S}_U) \text{,} \tag{8}$$

where, $U = \det\mathbf{S}_R / \det\mathbf{S}_U$. Note that in the single equation case the *LM* statistic reduces to TR^2(i.e. the number of observations times the *uncentered* coefficient of determination from the auxiliary regression). The above statistics are all asymptotically $\chi^2(p)$ distributed under the null hypothesis, where $p = Gn^2$ is the number of restrictions imposed by H_0. One simple small sample correction is to replace T by $\Delta = T - (m + Gn)$, the degrees of freedom in the

---

[1] It is obviously not possible to use $\hat{Y}_t$ in the auxiliary regression, since it is merely a linear combination of the exogenous variables.

[2] Ramsey (1969) suggested using a value of three to maximise power in common single equation situations, but it is not obvious that this result will carry over to systems of equations, where the loss of degrees of freedom can be considerable.

equations of the auxiliary regression. The *corrected* statistics are thus given by $WC = (\Delta/T)W$, $LRC = (\Delta/T)LR$ and $LMC = (\Delta/T)LM$, which have the same asymptotic distribution as given above.

Another more sophisticated approximation is that given by theorem (8.6.2) in Anderson (1958, p. 208). This uses an Edgeworth expansion, and if we choose the simplest form (which is accurate to the order T^{-2}) this corrected LR statistic is given by

$$LRE = \Delta_E \ln U, \tag{9}$$

where $\Delta_E = \Delta + ½[n(G-1)-1]$. This is also asymptotically $\chi^2(p)$ distributed under the null hypothesis. Note that when $G=1$, the difference between *LRC* and *LRE* is merely that the numerator in the correction is Δ in the first case and $\Delta - ½$ in the second.

A final approximation which is given by Rao (1973, p. 556), namely

$$RAO = (q/p)(U^{1/s} - 1), \tag{10}$$

where p and Δ_E are defined as above, $r = (p/2) - 1$, $q = \Delta_E s - r$, and

$$s = \sqrt{\frac{p^2 - 4}{n^2(G^2+1) - 5}}. \tag{11}$$

RAO is approximately distributed as $F(p,q)$ under the null hypothesis, and reduces to the standard F statistic when $n=1$. Note also that Rao's approximation reduces to Anderson's theorem 8.5.4 when $n=2$, a result that yields an exact distribution when applied to primary regressions.

3 Factors That Affect the Small Sample Properties of the RESET Test

A number of factors obviously can affect the size of the RESET tests, SE and SM have investigated these factors systematically, and we therefore follow their line of investigation. The number of equations (n), the sample size (T), degrees of freedom (Δ) and the order of the restrictions (G) are four such factors. The power of the tests will also be affected by the size and form of $Z_t \Gamma$ in (3). In this paper we will also study the consequences of varying n and Δ, while T is chosen so as to give compatible values of Δ for different models ($T = \Delta + m + Gn$). We will also mainly concentrate on the case where $G = 1$.

A number of other factors can also affect the properties of the RESET tests, for example the distributions of $\boldsymbol{X}_t$, and ε_t, and the values of B. In the rest of this section we will consider these in some more details. Regarding the distribution of the exogenous variables, a number of pilot studies, in SE, have been performed by SE and showed emphatically that trending the exogenous variables had no effect at all on either size or power. This factor has therefore not been varied in what follows.[3] In our Monte Carlo experiment, however, the exogenous variables have been obtained using the following generating processes,

$$x_{tj} = \alpha x_{t-1,j} + v_{tj}\,, \qquad j = 1, \ldots, \text{m-1}, \text{ and } t = 1, \ldots, \text{T}, \tag{12}$$

where and v_t is a multivariate normal white noise process with covariance matrix Σ_v. In our Monte Carlo study we have included a constant term among the exogenous variables, so that (12) has only been applied to the remaining $m-1$ variables.

The power of the tests will also be affected by $\mathbf{Z}_t\Gamma$ in (3). Intuitively, the power of the test ought to increase with an increase in the omitted portion of the regression. In other words, an increase in the absolute value of Γ should imply an increase in the seriousness of the misspecification which is caused by using (1) instead of (3). Accordingly, we would expect the power of the RESET test to increase with Γ. The problem is deciding how large a value of Γ is needed to signify "serious" misspecification.

SE found that good measure of misspecification is given by the relative increase of goodness-of-fit that is achieved by going from the incorrect model under the null (1) to the correct model under the alternative (3), *i.e.*,

$$R_D^2 = \frac{R_1^2 - R_0^2}{1 - R_0^2} \tag{13}$$

where R_0^2 and R_1^2 are the theoretical R^2 measures from the null and alternative models. The reasoning behind this choice of misspecification measure, and the relationships that exist between goodness-of-fit and the other parameters of the model, are explored in the Appendix. An advantage of using R_D^2 as a measure of misspecification is that it is bounded between zero (no misspecification) and one (a perfect alternative).

---

[3] SE have discovered that serial dependence in x had no practicable effect on either the size or power of the RESET tests.

The power of the test will also depend on the joint distribution of the included and omitted variables. If this is a joint normal distribution, then the regression of the omitted variables on the included variables is exactly linear, and no loss of fit will occur through the exclusion of the omitted variables. The RESET test will have zero power in such circumstances, even though the parameter estimates will be biased unless the omitted variable is also uncorrelated with the included variables. If the omitted variables are non-normal, then their conditional means can be nonlinear in the included variables, and the RESET test can have power. The strength of the power can well depend, however, on the correlation between the omitted variables and the proxy variables used in the auxiliary regression. In this paper, as in SE and SM, we concentrate on an omitted variable that is the square of one of the (normally distributed) included variables.

4 The Monte Carlo Experiment

The small sample properties of the systemwise RESET tests, when the residuals are generated from a non-normal distributions are unknown. Therefore, it is important to examine whether the actual behaviour of these tests is adequately approximated by asymptotic theory. In the absence of exact results, it is necessary to investigate the finite sample performance of the statistics by means of simulation experiments. When investigating the properties of a classical test procedure, two aspects are of prime importance. Firstly, we wish to see if the *actual size* of the test (*i.e.*, the probability of rejecting the null when it is true) is close to the *nominal size* (used to calculate the critical it is values). Given that actual size is a reasonable approximation to the nominal size, we then wish to investigate the *actual power* of the test (*i.e.*, the probability of rejecting the null when false) for a number of different alternative hypotheses. When comparing different tests we will therefore prefer those whose (a) actual size lies close to the nominal size and (b) have greatest power.

The primary interest lies in the analysis of systemwise tests, and thus the number of equations to be estimated is of central importance. As the number of equations grows the computation time becomes longer, and SE took a system with ten equations as our largest model when considering the size of the tests.

Another prime factor that affects the performance of tests is the number of observations. We, as in SE, have chosen to hold the number of degrees of freedom, Δ, constant between models of different sizes, since this allows a fair comparison. If the number of observations, T, was held constant then tests in models with a large number of equations would automatically perform more poorly, simply due to the reduced degrees of freedom (a new predictor is included for each equation in the system). We have investigated samples typical for annual and quarterly consumption models, using degrees of freedom 15, 25, 45 and 75. This is equivalent to sample sizes of between 20 and 110 observations.

Various values of R_0^2 were chosen to represent different explanatory powers under the null, with a greater variation in small models. The distribution of the exogenous variables was varied to account for a typical property of economic time series, *i.e.,* that they are trended and/or autocorrelated. As stated above, trending had no effect at all on the properties of the RESET test, and is therefore not reported here. We used different values of R_D^2 $(=0,0.1,\dots,0.9)$ to indicate different degrees of misspecification when calculating the power functions of the tests. For each replication of the experiment we generated a set of data, estimated the model using that data set, and computed the test statistics. These estimated statistics were then compared with the appropriate asymptotic critical values for the nominal size under question.

However, since we are most interested in the behaviour of the distributions in the tails, only results using the conventional 5% significance level have been analysed. A summary of the design we have used can be found in Tables 1 and 2. To judge the reasonability of the results, we require that the estimated size of the test should lay between the approximate 95% confidence intervals for the actual size presented in Table 3. Letting the number of replications per model be 10,000 seems therefore to be sufficient when estimating size, while 2,000 should suffice for the estimation of the power function.

Table 1. Values of Factors that Vary for Different Models - Size Calculations

Factor	Symbol	Design			
Number of equations	n	1	2	3, 5, 7	10
Degrees of freedom	Δ	15, 25, 45, 75			
Nominal size	π_0	1%, 5%, 10%, 20%			
Goodness-of-fit in null	R_0^2	.1, .3, .5, .7, .9	.3, .5, .7	.3, .7	.3 .7
AR parameter for X	α	0, .5, .9	0, .5		
Correlation (X_i, X_j)	ρ_x	0, .5, .9	0, .5		

Table 2. Values of Factors that Vary for Different Models - Power Calculations

Factor	Symbol	Design
Number of equations	n	1, 2, 3, 5, 7, 10
Degrees of freedom	Δ	15, 25, 45, 75
Nominal size	π_0	1%, 5%, 10%, 20%
Goodness-of-fit in null	R_0^2	.3, .5, .7
Relative difference in R^2	R_D^2	0, .1, .2, .3, .4, .5, .6, .7, .8, .9
AR parameter for X	α	0, .5
Correlation (η,z)	$\rho_{\eta z}$	.1, .3, .5, .7, .9

z is the omitted variable (the square of x_1) and η is the square of the conditional expected value of y.

Table 3. Approximate 95% Confidence Intervals for Actual Size

$\pi_0 \cdot N$	2000	10000
1%	±0.44	±0.20
5%	±0.97	±0.44
10%	±1.34	±0.60
20%	±1.79	±0.80

5 Analysis of the Size of the RESET Tests.

In this section we present the results of the main dominating effects of our Monte Carlo experiment concerning the size of the RESET tests. When determining the manner of presentation, some account has to be taken to the results obtained. We find, as in SE and SM, that the number of equations in the system (n) and the degrees of freedom (Δ) have a dominating effect on the performances of the tests. Moreover, we find that distribution of the error terms has also a very significant effect on the performances of the tests.

Some important results regarding the different variants of the RESET test are presented in following Table SE 4 (i.e., Table 4 in the SE study). Note that changing the factors we have held constant in this table (goodness-of-fit, multicollinearity and autocorrelation in X) did not change the conclusions in any way.

Table SE 4. Estimated Size for the Alternative RESET Tests at 5% Nominal Size.

	No. of Equations (n)						No. of Equations (n)					
	RAO						LRE					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.047	.048	.047	.050	.049	.058	.047	.048	.048	.062	.078	.182
25	.049	.047	.053	.048	.051	.048	.049	.047	.054	.051	.060	.078
45	.051	.051	.052	.049	.049	.048	.051	.051	.053	.050	.053	.055
75	.049	.050	.050	.054	.053	.054	.049	.050	.050	.054	.054	.056
	LRT-C						Wald-C					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.051	.056	.058	.082	.110	.249	.069	.120	.193	.504	.841	.998
25	.052	.051	.060	.061	.075	.103	.062	.085	.132	.279	.559	.921
45	.052	.053	.057	.054	.059	.065	.057	.074	.096	.162	.291	.602
75	.050	.051	.052	.058	.058	.062	.054	.062	.074	.109	.167	.339
	LRT						Wald					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.086	.164	.298	.756	.985	1.00	.101	.238	.457	.925	.999	1.00
25	.072	.107	.186	.468	.842	.999	.081	.150	.293	.708	.972	1.00
45	.062	.083	.116	.254	.500	.906	.067	.102	.165	.410	.760	.993
75	.058	.069	.086	.150	.284	.627	.060	.079	.113	.234	.469	.872
	LM						LM-C					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.071	.087	.112	.285	.608	.970	.037	.012	.003	0.00	0.00	0.00
25	.063	.069	.090	.162	.353	.763	.042	.025	.011	.001	0.00	0.00
45	.058	.062	.073	.105	.187	.419	.047	.036	.026	.008	.001	0.00
75	.054	.058	.062	.083	.118	.241	.048	.042	.035	.021	.009	.002

Source: Shukur and Edgerton (1997, Table 4). In this table $R_0^2 = 0.7$, $\rho_x = 0.5$ and $\alpha = 0.0$. The shading indicates bad performance, *i.e.*, when the results lie outside the approximate 95% confidence interval for actual size.

The most obvious result is that the *RAO* test is clearly superior to all the other alternatives, with only one result (of 24) outside the 95% confidence interval. The next best test is the Edgeworth adjusted likelihood ratio test, *LRE*, which performs well in systems up to 3 equations, but then gradually begins to deteriorate in small samples. In the case of 10 equations, the test performs badly for all sample sizes. The simpler likelihood ratio correction, *LRC*, perform well in small systems of 1 or 2 equations, but otherwise quite poorly. All the other tests perform poorly, only occasionally giving results that lie within the given confidence interval. Note that in all cases, except for the corrected *LM* test, we find that the tests tend to reject too often. For large systems and small samples the uncorrected *Wald* and *LRT* tests are rejecting 100% of the time!.

The result that the *RAO* test is best agrees with what is found by Edgerton and Shukur (1998) and Shukur (1998) for the Breusch-Godfrey autocorrelation test. In these cases all tests performed badly in large systems, and the difference between the *RAO* and *LRE* tests was smaller. The superiority of the Rao's *F*-test is, in other words, much more marked for RESET than for the Breusch-Godfrey test.

Because of this fact and to save space, we only present results for the *RAO* test in Table 5 when the error terms follow normal and t-distribution with different degrees of freedom in systems ranging from one to ten equations. To make a fair comparison between our results with those found by SE we present, in Table 5, the results for the situations when the $R^2 = 0.7$, $\rho_X = 0.5$, $\alpha = 0.0$, and $\pi_0 = 5\%$. Full results with other combinations are available from the author upon request.

In our analysis we have identified one factor which shows to have significant effect on the performance of the RESET tests, that is the heaviness of the tails of the error terms. When the error terms are generated from the $t_{(1)}$-distribution, even the best *RAO* test performs badly in all sample and system sizes. The effect is more dramatic in large system of equations which in this case the tests reject about 90% of the times under the null hypothesis. This bad effect becomes less and less with higher degrees of freedom in the t-distribution. In models with one equation, we achieve the same properties as in the normal case when the errors are generating from the $t_{(5)}$-distribution. In systems with more than 7 equations, we need to generate from, at least, the $t_{(25)}$-distribution to achieve the same properties as when the errors are normally distributed.

Table 5. Estimated Size for the RAO test with Alternative Distributions in the Error Terms at 5% Nominal Size.

	No. of Equations (n)						No. of Equations (n)					
	RAO - Normal						RAO – $t_{(25)}$					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.047	.048	.047	.050	.049	.058	.047	.052	.051	.050	.054	.056
25	.049	.047	.053	.048	.051	.048	.054	.055	.053	.051	.050	.054
45	.051	.051	.052	.049	.049	.048	.051	.050	.054	.055	.050	.053
75	.049	.050	.050	.054	.053	.054	.048	.050	.051	.050	.050	.051
	RAO – $t_{(15)}$						RAO – $t_{(10)}$					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.051	.049	.052	.052	.053	.059	.052	.050	.o53	.058	.058	.062
25	.051	.054	.053	.054	.057	.056	.057	.055	.053	.056	.056	.058
45	.051	.053	.053	.052	.055	.053	.050	.053	.059	.054	.058	.056
75	.050	.051	.055	.056	.052	.059	.052	.055	.051	.051	.058	.056
	RAO – $t_{(7)}$						RAO – $t_{(5)}$					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.050	.052	.060	.058	.061	.062	.048	.059	.059	.069	.070	.073
25	.049	.049	.057	.056	.063	.065	.053	.059	.059	.066	.072	.076
45	.052	.057	.057	.054	.058	.064	.050	.058	.064	.070	.071	.081
75	.049	.054	.055	.051	.063	.063	.051	.057	.058	.064	.069	.073
	RAO – $t_{(2)}$						RAO – $t_{(1)}$					
Δ	1	2	3	5	7	10	1	2	3	5	7	10
15	.066	.099	.128	.195	.249	.321	.119	.221	.325	.567	.761	.919
25	.058	.093	.125	.205	.295	.427	.108	.213	.334	.601	.820	.964
45	.056	.091	.123	.207	.314	.467	.101	.199	.333	.610	.840	.980
75	.059	.086	.118	.191	.290	.461	.096	.196	.324	.610	.843	.9796

In this table $R_0^2 = 0.7$, $\rho_x = 0.5$ and $\alpha = 0.0$. The shading indicates bad performance, *i.e.*, when the results lie outside the approximate 95% confidence interval for actual size.

In Table 6, however, we present results of our experiment when we use the bootstrap critical values instead. The results will be compared with those found be the SM. Our primary results, as in the SM, reveal that the LM and Wald tests give results identical to their corrected correspondents (i.e., LMC, and WC).

Table 6. Estimated Size for the RAO test with Alternative Distributions in the Error Terms at 5% Nominal Size (using the bootstrap critical values).

	No. of Equations (n)											
Normal	Wald = Wald-C				RAO = LRE = LRT = LRT-C				LM = LM-C			
Δ	1	3	7	10	1	3	7	10	1	3	7	10
15	.049	.051	.050	.045	.049	.053	.046	.042	.049	.051	.047	.047
25	.050	.051	.047	.045	.050	.051	.048	.045	.050	.052	.049	.050
45	.051	.050	.055	.047	.051	.050	.054	.049	.051	.048	.054	.050
75	.047	.046	.049	.047	.047	.046	.050	.048	.047	.047	.050	.045
$t_{(10)}$												
Δ	1	3	7	10	1	3	7	10	1	3	7	10
15	.051	.052	.047	.044	.051	.052	.046	.045	.051	.051	.048	.046
25	.054	.050	.051	.050	.054	.050	.048	.050	.054	.050	.051	.049
45	.051	.050	.053	.050	.051	.048	.054	.049	.051	.048	.053	.051
75	.053	.049	.048	.050	.053	.049	.050	.050	.053	.049	.048	.052
$t_{(7)}$												
Δ	1	3	7	10	1	3	7	10	1	3	7	10
15	.053	.053	.048	.050	.053	.053	.053	.049	.053	.054	.051	.048
25	.047	.052	.051	.049	.047	.052	.052	.050	.047	.053	.055	.053
45	.054	.052	.054	.051	.054	.052	.054	.050	.054	.052	.056	.050
75	.048	.054	.051	.051	.048	.054	.049	.051	.048	.053	.050	.051
$t_{(5)}$												
Δ	1	3	7	10	1	3	7	10	1	3	7	10
15	.052	.054	.053	.053	.052	.054	.057	.054	.052	.053	.060	.054
25	.050	.057	.058	.055	.050	.058	.059	.055	.050	.059	.058	.059
45	.052	.058	.058	.061	.052	.058	.059	.062	.052	.058	.059	.060
75	.050	.056	.059	.058	.050	.056	.059	.059	.050	.056	.057	.059
$t_{(2)}$												
Δ	1	3	7	10	1	3	7	10	1	3	7	10
15	.060	.099	.164	.151	.060	.102	.186	.227	.060	.097	.166	.211
25	.056	.095	.200	.286	.056	.095	.196	.301	.056	.093	.176	.259
45	.055	.085	.184	.290	.055	.083	.178	.285	.055	.081	.169	.257
75	.052	.078	.159	.256	.052	.078	.156	.250	.052	.077	.151	.237
$t_{(1)}$												
Δ	1	3	7	10	1	3	7	10	1	3	7	10
15	.104	.221	.519	.572	.104	.244	.592	.802	.104	.209	.558	.785
25	.098	.194	.566	.809	.098	.195	.603	.856	.098	.192	.587	.836
45	.090	.176	.521	.803	.090	.176	.549	.835	.090	.177	.553	.837
75	.078	.165	.472	.792	.078	.164	.495	.790	.078	.165	.508	.793

In this table $R_0^2 = 0.7$, $\rho_X = 0.5$ and $\alpha = 0.0$. The shading indicates bad performance, *i.e.*, when the results lie outside the approximate 95% confidence interval for actual size.

We also find that all the LR tests (including the RAO) lead to identical results. Moreover, for a single equation, we find that all the eight test methods yield the same results. Note that, when using the bootstrap critical values and to achieve the same properties as when the errors are normally distributed, the errors only need to be generated from the t-distribution with seven degrees of freedom. This means that, in situations when the error terms are rather heavily tailed (not heavier than $t_{(7)}$), the RESET test performs better when using the bootstrap critical values instead of the critical values from the F- and χ^2 -distributions.

6 Analysis of the Power of the RESET tests

In this section we discuss the most interesting results of our Monte Carlo experiment, concerning the power of the various versions of the RESET test. The effect of the non-normally distributed error terms on the power functions of different versions of the RESET test is analysed in systems ranging from one to ten equations. The power function was estimated by calculating the rejection frequencies in 2,000 replications using different values of the relative differences in goodness-of-fit, R_D^2.

In the SE study, the authors shown that the corrected LR tests are the only ones that accurately estimate the size in a wide variety of situations, and in particular the RAO tests were shown to be superior in all situations. SE therefore only present results for the RAO test. On the other hand, since all tests perform well regarding the size when using the bootstrap critical values, SM compare the power functions of the three test groups (Wald, LR and LM). SM find that, for single equation, the power functions for all the tests are identical. This means that in single equation, the eight tests reduces to one and that one can present results from any one of them. In systems with more than one equation the results differ somewhat between the three test groups. All tests have shown satisfactory power functions specially in small systems and large samples, and that the differences between the alternative RESET tests are very minimal. Note that the RAO test performs well in both studies SE and SM. Accordingly, to show the effect of the non-normally distributed error terms and to facilitate comparison between the three papers, we will only present results for the RAO test in this study.

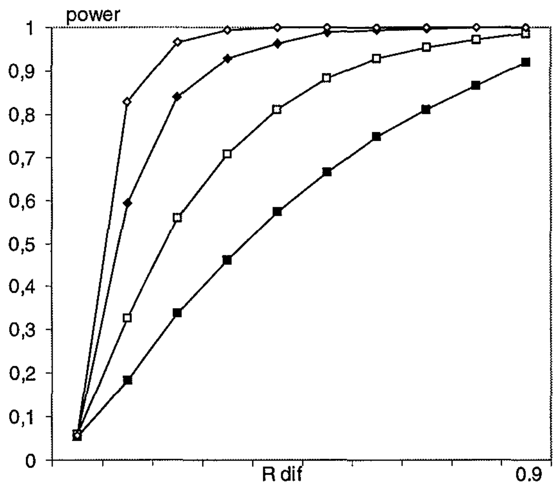
The number of equations (n), degrees of freedom (Δ) and the distribution of the error terms had a considerable effect on both size and power (except for the case when using the bootstrap critical values together with normally distributed errors, in this case the tests perform satisfactorily in almost all situations). As in the case of the size, changes in the autocorrelation between the exogenous variables α and the goodness-of-fit in the null (R_0^2) did not produce any noticeable effects on the estimated size or the power, and will not be shown in the figures. Finally, the correlation between the included and omitted variables ($\rho_{\eta z}$) also seems to affect the power function. The greater the misspecification, and the better the RESET proxy mirrors the omitted variable, the greater the power of the tests.

In SE only results for the *RAO* test have been presented, which have shown to be very similar to those results for the other tests groups in SM. Hence, to make our results comparable to those of SE and SM, we use the same combinations as in SE and SM and fix the autocorrelation in the exogenous variables ($\alpha = 0$), the goodness-of-fit in the null ($R_0^2 = 0.7$) and the correlation between the included and omitted variables ($\rho_{\eta z} = 0.5$).

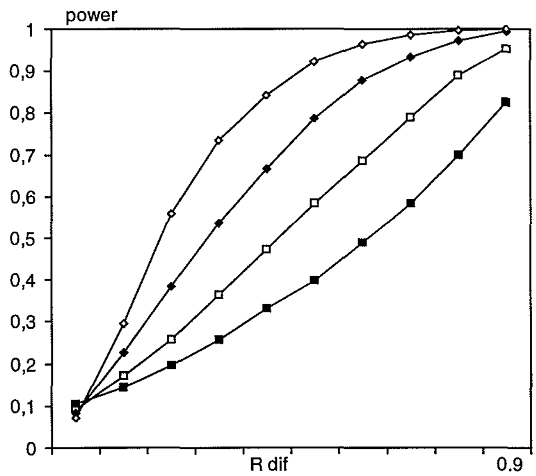
In Figure 1 we show the power functions of the *RAO* test, at a nominal size of 5% for different degrees of freedom (Δ) and for systems with different numbers of equations (n). The power functions have also been calculated for other values, but since the patterns obtained are essentially the same they are excluded to save space.

The power of the RESET test did, as expected, depend on the degree of misspecification (R_D^2). Note also how the power functions become flatter as the number of equations increases and Δ decreases, *i.e.,* the *RAO* test becomes worse and worse. The distribution of the error terms has also shown to affect the properties of the RESET test especially in large systems and when the t-distributions have low degrees of freedom. Note that when choosing the t-distribution with seven degrees of freedom, the results are almost identical to the situations when the errors are generated from the normal distribution. In order to show the effect of the fat tails on the properties of the RESET test, we used $t_{(2)}$-distribution to generate the error terms instead of the $t_{(1)}$-distribution, since the later have shown to have extremely biased results regarding the size of the test.

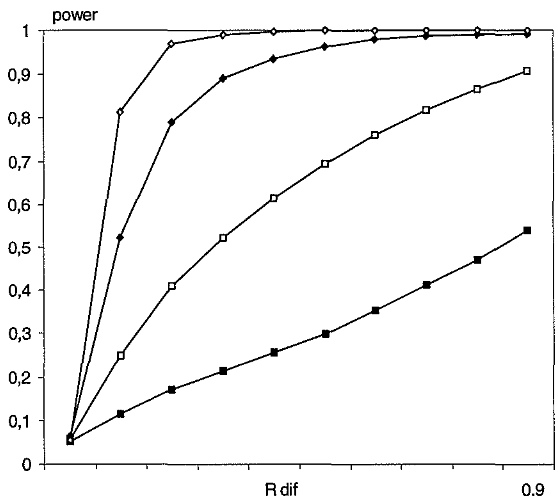
Figure 1. The Power Function of the RAO Test Using the Bootstrap Critical values, when the Errors Follow $t_{(7)}$ $t_{(2)}$ Distributions.



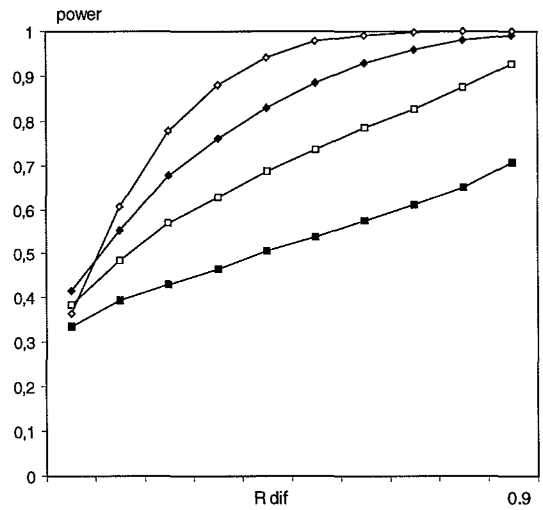
3 equations, $t_{(7)}$



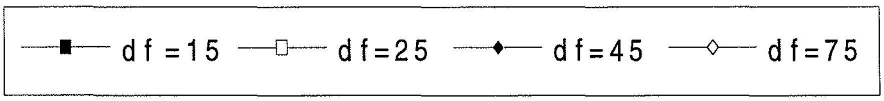
3 equations, $t_{(2)}$



10 equations, $t_{(7)}$



10 equations, $t_{(2)}$



Note that the power functions have all been estimated using the nominal size, even in those situations where we have found that it does not correspond to the actual size. The results must therefore be interpreted as rejection rates at nominal significance levels, not as true power functions. One could, of course, calculate and present the size-corrected power functions which give a more correct information about the power of the tests. However, there is one drawback in using this method, namely that the reader can get a good idea about the real power but a misleading idea about the performance of the size (when corrected). For this reason, we decide to use the rejection rates at nominal significance levels and leave the reader to make the inferential statements regarding the performances of both the size and power.

7 Conclusions

In this paper we have studied the properties of systemwise generalisations of Ramsey's RESET test for misspecification errors when the error terms follow a normal distribution and t-distribution with different degrees of freedom. The degrees of freedom of the t-distribution have been chosen so that to cover a varying amount of fatness in the tails of the errors, which could, of course, affect the performances of the test.

We have constructed Wald, Lagrange Multiplier and Likelihood Ratio tests that are applied to auxiliary regression systems. Various degrees-of-freedom corrections have been investigated, in particular the commonly used simple replacement of the number of observations (T) by the degrees-of-freedom (Δ) and, for the LR test, the Edgeworth correction test and Rao's multivariate F-test.

The investigation has been carried out using Monte Carlo simulations. A large number of models were investigated, where the number of equations, degrees of freedom, error variance and stochastic properties of the exogenous variables have been varied. For each model we have performed 10,000 replications and studied four different nominal sizes. The power properties have been investigated using 2,000 replications per model, where in addition to the properties mentioned above the degree of misspecification (measured as the relative differences in the explanatory power between the null and true models) and the correlation between the omitted and included variables have also varied.

Our analysis has revealed that two factors are important when determining the accuracy of the RESET-tests' nominal size (except in case when using the bootstrap critical values together with normally distributed errors), namely the number of equations and number of observations. When using a normally distributed error terms, the performance of the tests (except for the *RAO* test) deteriorates as the first factor increases and the second factor decreases. Moreover, we found that the fatness of the tails of the error terms has also a considerable effect on the test properties especially in large systems.

ACKNOWLEDGMENTS

The author is grateful to professor Kasra Afsarinejad for a careful reading of the manuscript and valuable comments. The author is also grateful to professor Marianne Frisén for encouragement and support.

References

Anderson, T. W. (1958): *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Edgerton, D. L. and G. Shukur (1998): "Testing Autocorrelation in a System Perspective," accepted for publication, to appear in *Econometrics Reviews* 1999.

Ramsey, J. B. (1969): "Test for Specification error in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society, Series B*. 31, 350-371.

Rao, C. R. (1973): *Linear Statistical Inference and Its Applications*, Second edition. New York: Wiley.

Shukur, G., (1998): "The Robustness of the Systemwise Breauch-Godfrey Autocorrelation Test for Non-normal Distributed Error Terms, " Department of Statistics, University of Göteborg, Sweden. Submitted for publication, *Communications in Statistics*.

Shukur, G., and D. L. Edgerton (1997): "The Small Sample Properties of the RESET Test as Applied to Systems of Equations," FD Thesis, Department of Statistics, Lunds University, Sweden.

Shukur, G., and P. Mantalos (1997): "Size and Power of the RESET Test as Applied to Systems of Equations : A Bootstrap Approach, " Department of Statistics, *Working paper* 1997:3, Lunds University, Sweden.

Research Report

1998:6	Dahlbom, U.:	Least squares estimates of regression functions with certain monotonicity and concavity/convexity restrictions.
1998:7	Dahlbom, U.:	Variance estimates based on knowledge of monotonicity and concavity properties.
1998:8	Grabarnik, P. & Särkkä, A.	Some interaction models for clustered point patterns.
1998:9	Afsarinejad, K. & Hedayat, S.:	Repeated measurement designs for models with self and mixed carryover effects.
1998:10	Hatemi-J, A. & Shukur, G.:	The causal nexus of government spending and revenue in Finland: A bootstrap approach.
1998:11	Shukur, G.:	The robustness of the systemwise Breauch--Godfrey autocorrelation test for non-normal distributed error terms.
1999:1	Andersson, E.:	On monotonicity and early warnings with applications in economics.
1999:2	Wessman, P.:	The surveillance of several processes with different change points.
1999:3	Andersson, E.:	Monotonicity aspects on seasonal adjustment.
1999:4	Andersson, E.:	Monotonicity restrictions used in a system of early warnings applied to monthly economic data.
1999.5	Mantalos. P. & Shukur, G.:	Testing for cointegrating relations- A bootstrap approach.