A MODIFICATION OF THE CUSUM TEST

Raimund ALT
Walter KRÄMER

Forschungsbericht/
Research Memorandum No. 229

März 1986

Summary

We show via Monte Carlo experiments that the power of the Brown-Durbin-Evans CUSUM test for the stability of regression relationships over time can be improved upon by using an alternative estimate for the variance of the disturbances in the regression.

1. INTRODUCTION

This paper examines the power of the Brown-Durbin-Evans (1975) CUSUM test for the stability of regression relationships over time. Though this procedure was initially proposed more as a diagnostic device rather than a formal significance test, it is often criticised for its alleged lack of power. Garbade (1977), for instance, who might well be responsible for this popular belief obtained very little power for the CUSUM test in his Monte Carlo experiments.

Below we demonstrate with the help of another series of Monte Carlo experiments that the poor performance of the CUSUM test in Garbade's study is specific to his particular experimental design, and that the power of the CUSUM test can further be improved upon by using a different estimate for the variance of the disturbances in the equation.

2. THE MODEL AND THE TEST

We consider the regression model

$$y_t = x_t'\beta_t + u_t \quad , \qquad t=1,\dots,T, \tag{1}$$

where y_t is the observation on the dependent variable at time t, x_t is a Kx1 vector of observations on the independent variables (which are assumed non-stochastic), β_t is a Kx1 vector of regression coefficients, and u_t is a stochastic disturbance. We further assume that these disturbances are independent and normally distributed with means zero and common variance σ^2. The vector of regression coefficients may however vary with time, as indicated by the subscript t. The null hypothesis under test is that the regression coefficients remain constant over time, i.e. $\beta_t = \beta$, $t = 1,\ldots,T$.

The CUSUM test by Brown, Durbin and Evans (1975) is based on recursive residuals, which for $t > K$ are defined as

$$\tilde{u}_t = (y_t - x_t'\hat{\beta}^{(t-1)})/f_t \; , \tag{2}$$

where $\hat{\beta}^{(t-1)}$ is the OLS-estimate for β from the first $(t-1)$ observations,

$$f_t = (1 + x_t'(X^{(t-1)'}X^{(t-1)})^{-1}x_t)^{1/2} \; , \tag{3}$$

and $X^{(t)} = [x_1, \ldots, x_t]'$.

We require here that $X^{(t)}$ has full column rank for $t = K$ (and thus for all $t > K$). The CUSUM test can easily be modified to accomodate rank deficiencies in some initial regressor matrices, a complication that will often arise in practice when there are dummy variables in the regression, but such complications shall not concern us here.

From (2) it is obvious that the recursive residuals are just standardised forecast errors, where the normalisation factor (3) ensures that under H_o the variance of the recursive residuals equals σ^2 for all $t = K + 1,\ldots,T$. In addition, the recursive residuals will then be independent and normally distributed.

The CUSUM test rejects the null hypothesis of parameter stability whenever the standardized cumulated sums of the recursive residuals

$$W^{(r)} = \frac{1}{\hat{\sigma}} \Sigma_{t=K+1}^{r} \tilde{u}_t \tag{4}$$

are too large, i.e. cross either the upper or lower critical line as given in Brown, Durbin and Evans (1975, p. 153).

The quantity $\hat{\sigma}$ in (4) denotes the estimated standard deviation of the regression disturbances, and will in the sequel be of major concern to us.

Brown, Durbin and Evans (1975, p. 153) suggest to determine $\hat{\sigma}$ in standard fashion from

$$\hat{\sigma}^2 = \hat{u}'\hat{u}/(T-K) \quad , \tag{5}$$

where $\hat{u} = (y^{(T)} - X^{(T)}\hat{\beta}^{(T)})$ is the vector of OLS regression residuals from the full sample. Both under H_o and under the alternative, the arithmetic mean $\bar{\hat{u}}$ will be zero (given that there is a constant in the regression), so it does not matter whether or not we subtract $\bar{\hat{u}}$ from u_t when estimating σ^2.

This is not true when we estimate σ^2 from the recursive residuals, which need not have a sample mean of zero. Harvey (1975) has therefore proposed to estimate σ by the square root of

$$\tilde{\sigma}^2 = \Sigma_{t=K+1}^{T} (\tilde{u}_t - \bar{\tilde{u}})^2/(T-K-1) \tag{6}$$

rather than by the square root of (5). This will not affect the theory behind the CUSUM test, since under H_o, both $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ are consistent estimators for σ^2, but will hopefully enhance its power.

Depending upon the type of structural change, one will under the alternative often have a very erratic pattern of OLS residuals, and thus, when using (5), a gross overestimation of σ^2. Ceteris paribus, any such overestimation of σ^2 will reduce the probability of a rejection of the null hypothesis, since it is obvious from (4) that the cumulated sums are then less likely to cross the critical lines.

On the other hand, the upward bias in $\tilde{\sigma}^2$ will in general be much smaller, since $\tilde{\tilde{u}}$ will on average be far away from zero for many types of structural change. To the extent therefore that the overestimation of $\hat{\sigma}^2$ is mitigated by $\tilde{\sigma}^2$, we will have more power for the CUSUM test.

3. THE LOCAL POWER OF THE CUSUM TEST

Assume that

$$\lim_{T\to\infty} \frac{1}{T} X^{(T)'}X^{(T)} = Q \qquad (7)$$

for some nonsingular KxK matrix Q. Whenever there is a constant in the regression, this implies that

$$\lim_{T\to\infty} \frac{1}{T} \Sigma_{t=1}^{T} x_t = q \qquad (8)$$

for some K-vector (mean regressor) q (when the constant is first in the regression, q is equal to the first column of Q). When there is no constant in the regression, impose (8) as an additional assumption.

Let g(z) $(0 \leq z \leq 1)$ be some K-dimensional function, and consider the following sequence of alternatives as $T\to\infty$:

$$\beta_t^{(T)} = \beta + \frac{1}{\sqrt{T}} g(t/T), \qquad (9)$$

Ploberger (1983) has shown that when g(z) is orthogonal to q for all z, we have for any significance level α that

$$\lim_{T\to\infty} P(\text{CUSUM test rejects } H_o) = \alpha, \tag{10}$$

i.e. there is no nontrivial local power for the CUSUM test. This holds irrespective of whether $\tilde{\sigma}^2$ or $\hat{\sigma}^2$ is used to estimate σ^2.

Ploberger's result has important implications for Garbade's (1977) Monte Carlo experiments, where K=1 and the independent variable is generated to follow a normal distribution with mean zero and variance 25 (the same design was also used by McCabe and Harrison, 1980). Here the mean regressor equals q = 0 and will thus be orthogonal to any structural change. In view of (10), it is therefore not surprising at all that the power of the CUSUM test turned out to be very weak.

An important special case is when there is just one structural shift $\Delta\beta$ at time $T^* = dT$, where d is some constant ($0 < d < 1$). Here, $g(z) = 0$ for $z < d$ and $g(z) = \Delta\beta$ for $z \geq d$. The CUSUM test will then have little power when $\Delta\beta$ is orthogonal to q.

4. THE FINITE SAMPLE NULL DISTRIBUTION

Table 1 reports the empirical rejection probabilities for both the standard and the modified CUSUM test, for various sample sizes T and significance level α equal to one, five and 10 percent. The model was

$$y_t = (-1)^t + 1 + u_t \quad , \tag{11}$$

i.e. K = 2 and $\beta = [1,1]'$. The particular form of the regressor matrix was chosen to ensure that (7) and (8) hold, and that both Q and q are well approximated by respectively $X^{(T)\prime}X^{(T)}/T$ and $\Sigma X_t/T$ in small samples. The disturbances u_t were generated as standard normal variates, using various NAG-library subroutines, and 1000 replications (trials, runs) were performed for any given T.

The tests were done with programs from the IAS-SYSTEM econometric software package (see Sonnberger and Krämer, 1984). The same artificial samples were used for both types of the CUSUM test.

Table 1 provides additional empirical evidence for the well known fact that the actual size of the standard CUSUM test is in small samples considerably smaller than the nominal significance level. To a lesser extent, this appears also be true for the modified CUSUM test, whose empirical rejection rates are consistently above those for the standard test, sometimes substantially so.

Both tests approach their nominal size as T increases. The ups and downs in the empirical rejection rates as T increases can be attributed to the sampling variability in the Monte Carlo results.

5. THE FINITE SAMPLE POWER

We keep the basic model as in (11), and confine ourselves to the empirically most relevant sample sizes T = 30, 60, 120. First we consider a single discrete jump in β at time $T^* = dT$, where d equals 0.3, 0.5 and 0.7, respectively. This corresponds to a structural change early, midway and late in the sample period. The shift in β is given by

$$\Delta\beta = \frac{b}{\sqrt{T}} [\sin \Psi, \cos \Psi]' . \qquad (12)$$

From (11) it is immediately seen that the mean regressor in our model equals $q = [0,1]'$, so Ψ in (12) is the angle between the mean regressor q and the structural shift $\Delta\beta$. We let Ψ take the values 0°, 36° and 90°. The intensity of the shift is $||\Delta\beta|| = |b|/\sqrt{T}$, which we let go down with increasing sample size in order to have the empirical rejection rates approximate the local power of the tests. 1000 experiments were performed for any given combination of T, b, d and Ψ, and both versions of the CUSUM test were applied to the same artificially generated data sets.

Table 2 reports the resulting empirical rejection rates, for a nominal significance level of $\alpha = 5$ %.

Several regularities stand out (ceteris paribus):
(i) the power increases for both tests as the intensity of the shift increases, except for $\Psi = 90^{\circ}$ (i.e. a shift in the slope only). Here, the increase in the intensity blows up both $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ more than the cumulated sums of recursive residuals, and thus reduces the power of the tests (ii) the power decreases as the angle between the structural shift and the mean regressor increases, confirming Ploberger (1983) (iii) the power increases as T increases, even though the intensity of the shift decreases with $1/\sqrt{T}$. This is most probably due to the increase in the true size of the tests, as shown in Table 1. (iv) the power decreases as the shift point moves towards the end of the sample period. This is not surprising, since then the CUSUM test have less time to react to any structural change.(v) the modified CUSUM test consistently outperformes the standard CUSUM test. The relative difference in power is highest for small sample sizes, and decreases as T increases, confirming the result in Ploberger (1983) that the local power of the test does not depend on the particular estimate for σ^2.

Table 3 reports the analogues results for two structural shifts, at times $T_1^* = 0.3\ T$ and $T_2^* = 0.7\ T$, respectively. The first shift is as in (12), whereas the second shift is either equal to $-2\Delta\beta$ or identical to the first shift. For $t > T_2^*$, the regression coefficients are thus either $\beta-\Delta\beta$ or $\beta+2\Delta\beta$. Under the heading of "rw", Table 3 also reports the empirical power when the regression coefficients follow a random walk. This means that for $t > 2$, the regression coefficients were generated according to

$$\beta_t = \beta_{t-1} + \eta_t \Delta\beta \ , \qquad (13)$$

where the η_t are nid(0,1), and $\Delta\beta$ is as in (12).

Table 3 confirms the results of Table 2, except that both variants of the test have no very little power for small sample size, and when the shifts are counteracting each other. The decrease in power in the face of an increase in the intensity of the shifts that now occurs also for $\Psi \neq 90^\circ$ is again due to a more than proportional increase in the bias of both $\hat{\sigma}^2$ and $\tilde{\sigma}^2$. These figures were also recorded for most experiments, but are not explicitely reproduced in the tables. The increase in both $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ when b increases is also responsible for the small response of the power for both tests to changes in b when the regression coefficients follow a random walk.

6. CONCLUSION

We conclude that the bad reputation of the CUSUM test is mostly due to the particular experimental design of previous Monte Carlo studies, and that the power of the test can considerably be improved by using an alternative estimate for the variance of the regression disturbances.

TABLE 1: Empirical rejection rates under H_o for the standard and modified CUSUM tests

	sample size T					
α	10	20	30	60	120	1000
			a) standard CUSUM test			
1 %	0	0	0.1	0.5	0.3	1.1
5 %	0	1.6	2.0	4.2	3.7	5.0
10 %	1.3	4.0	6.1	8.4	8.1	10.7
			b) modified CUSUM test			
1 %	2.1	0.7	0.6	1.0	0.6	1.2
5 %	5.5	3.7	4.7	4.9	3.9	5.0
10 %	9.2	5.9	8.1	9.6	8.4	10.7

TABLE 2: Empirical rejection rates for a single structural shift

	b	T=30			T=60			T=120		
		0^o	36^o	90^o	0^o	36^o	90^o	0^o	36^o	90^o
					a) standard CUSUM test					
d=0.3	4.8	16.1	10.6	1.4	21.7	12.6	2.6	22.2	15.4	2.5
	7.2	40.8	19.8	0.6	47.4	31.4	1.5	53.5	32.8	2.8
	9.6	62.3	29.7	0.2	73.1	47.2	1.0	80.3	56.2	2.0
	12.0	81.1	48.0	0.0	92.1	65.6	0.8	94.0	78.6	0.8
d=0.5	4.8	10.1	5.6	1.2	14.4	8.6	2.2	15.8	8.9	2.7
	7.2	22.0	9.4	0.6	32.0	16.9	1.3	40.6	20.6	2.4
	9.6	38.5	17.5	0.2	56.8	32.1	0.7	65.5	41.7	2.3
	12.0	55.4	25.0	0.0	78.4	45.0	0.4	88.5	61.1	1.5
d=0.7	4.8	3.2	3.1	1.0	4.4	4.9	2.2	7.0	5.1	3.3
	7.2	4.2	3.0	0.6	9.5	7.7	1.5	15.1	9.7	2.4
	9.6	9.4	3.6	0.2	19.0	10.5	1.1	28.8	12.9	1.9
	12.0	14.0	4.5	0.0	31.2	13.8	0.3	47.8	25.4	1.2
					b) modified CUSUM test					
d=0.3	4.8	26.2	18.1	2.6	25.9	16.0	3.3	23.9	16.7	3.0
	7.2	55.9	33.1	1.4	54.1	36.7	1.8	57.0	35.9	3.0
	9.6	78.4	48.2	0.4	78.7	55.8	1.2	82.7	59.6	2.2
	12.0	92.2	67.5	0.1	94.3	72.4	0.9	95.0	81.1	1.1
d=0.5	4.8	17.2	12.3	2.5	19.3	11.4	3.1	17.8	9.7	3.2
	7.2	35.7	19.5	1.2	37.1	21.3	1.8	43.4	24.0	2.6
	9.6	60.3	32.5	0.8	65.3	39.4	1.0	69.1	45.1	2.6
	12.0	76.4	44.6	0.1	85.8	54.4	0.5	91.1	66.9	1.9
d=0.7	4.8	6.7	6.6	2.8	5.9	6.3	3.1	8.2	5.7	3.4
	7.2	10.9	6.7	1.5	13.7	10.7	1.8	17.5	11.6	2.6
	9.6	21.2	11.6	0.5	26.5	15.3	1.2	33.6	16.0	2.1
	12.0	31.9	16.2	0.3	39.7	19.1	0.7	51.9	29.0	1.4

TABLE 3: Empirical rejection rates for several structural shifts

	b	T=30			T=60			T=120		
		0^o	36^o	90^o	0^o	36^o	90^o	0^o	36^o	90^o
					a) standard CUSUM test					
2nd sh. =$\Delta\beta$	4.8	37.3	16.1	0.4	48.6	27.1	1.6	55.1	32.5	2.6
	7.2	75.0	36.0	0.1	88.6	57.1	0.4	91.2	71.7	1.9
	9.6	94.8	57.8	0.0	98.9	85.0	0.0	99.8	92.6	0.8
	12.0	99.3	74.0	0.0	100.0	96.5	0.0	100.0	99.6	0.1
2nd sh. =$-2\Delta\beta$	4.8	3.5	2.8	0.3	9.3	7.2	1.3	13.6	9.3	2.1
	7.2	3.8	1.8	0.0	13.8	8.1	0.2	28.3	15.9	1.3
	9.6	3.4	0.7	0.0	19.3	8.2	0.1	42.2	22.2	0.8
	12.0	2.2	0.4	0.0	25.4	7.2	0.0	55.6	30.6	0.2
rw	4.8	56.5	38.3	0.0	80.9	67.6	0.0	94.7	87.5	0.2
	7.2	64.3	47.1	0.0	85.5	74.3	0.0	97.0	91.7	0.1
	9.6	64.1	49.4	0.0	88.6	73.9	0.0	97.3	92.6	0.0
	12.0	68.2	49.8	0.0	88.1	76.4	0.0	97.6	91.5	0.0
					b) modified CUSUM test					
2nd sh. =$\Delta\beta$	4.8	55.6	30.2	0.8	56.6	34.1	2.5	57.9	36.9	3.2
	7.2	90.1	60.3	0.1	93.0	66.8	0.8	93.0	75.0	2.0
	9.6	98.4	80.7	0.1	99.5	89.7	0.1	99.8	94.1	0.8
	12.0	100.0	91.8	0.0	100.0	98.5	0.0	100.0	99.7	0.1
2nd sh. =$-2\Delta\beta$	4.8	4.3	3.5	1.2	10.0	7.8	1.6	14.0	9.9	2.5
	7.2	4.2	2.0	0.0	14.1	8.5	0.2	29.0	16.7	1.4
	9.6	3.5	0.8	0.0	20.2	8.5	0.1	43.5	22.7	0.9
	12.0	2.2	0.4	0.0	25.8	7.2	0.1	57.8	31.3	0.3
rw	4.8	62.1	48.3	0.2	83.4	70.4	0.0	95.1	88.4	0.2
	7.2	70.2	55.4	0.0	88.0	77.1	0.0	97.3	92.2	0.1
	9.6	70.3	58.6	0.0	90.1	76.7	0.0	97.4	93.1	0.0
	12.0	73.7	58.5	0.0	89.3	78.9	0.0	97.9	92.3	0.0

REFERENCES

Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationship over time. J.R. Statist.Soc. B, 37, 149-163.

Garbade, K. (1977). Two methods for examining the stability of regression coefficients, J.Amer.Statist.Ass., 72, 54-63.

Harvey, A.C.(1975). Discussion of the paper by Brown, Durbin and Evans, J.R.Statist.Soc. B, 37, 179-180.

McCabe, B.P.M., and Harrison, M.J. (1980). Testing the constancy of regression relationships over time using least squares residuals, Applied Statistics, 29, 142-148.

Ploberger, W. (1983). Testing the constancy of parameters in linear models, paper given at the Econometric Society European Meeting, Pisa.

Sonnberger, H. and Krämer, W. (1984). The IAS-SYSTEM: an econometric software package for the analysis of time series data, Statistical Software Newsletter, 10, 136-139.