

Web based Supplementary Material for *Testing
with a nuisance parameter present only under
the alternative: a score-based approach with
application to segmented modelling*

Vito M.R. Muggeo*

Web Appendix: Testing for a breakpoint in Generalized Linear Models

This supplementary material includes some additional simulation results concerning performance of the proposed Score approach for testing for $H_0 : \delta = 0$ in generalized linear segmented regressions. Namely for response variable Y and quantitative covariate x , the regression equation is $E[Y|x_i] = \mu_i = g^{-1}(\eta_i)$ where $g(\cdot)$ is the known link function and $\eta_i = \theta_0 + \theta_1 x_i + \delta(x_i - \psi)_+$ is the linear predictor. As outlined in the paper, the proposed score statistic s_0 easily applies to regression models with non-Gaussian responses. In this supplementary note we consider the popular Binomial logistic and Poisson log linear models. The score statistic and its variance are

$$\dot{\ell}_0 = \bar{\varphi}^T(I_n - A)y \quad \text{var}(\dot{\ell}_0) = \bar{\varphi}^T(I_n - A)W(I_n - A)\bar{\varphi},$$

where y is the observed response vector, W is the diagonal weight matrix, and the hat matrix is

$$A = W^{1/2}X(X^TWX)^{-1}X^TW^{1/2}.$$

To compute $\dot{\ell}_0$ and $\text{var}(\dot{\ell}_0)$, both A and W refer to the null model and thus estimation under H_1 is not requested.

The proposed score test statistic is simply

$$s_0 = \dot{\ell}_0^2 / \text{var}(\dot{\ell}_0) \xrightarrow{d} \chi_1^2.$$

under H_0 .

In the following we report some simulation results comparing s_0 with the approach of Davies (1987) via his formula (3.4) using the likelihood ratio statistic. We emphasize that s_0 needs a single fit under the null hypothesis, while

*Università di Palermo. Email: vito.muggeo@unipa.it

the Davies test additionally requires several fits at different fixed values of the nuisance ψ .

Binomial logit regression

To compare the Score and the Davies tests we consider different scenarios: $y_i \sim \text{Bin}((1 + \exp\{-\eta_i\})^{-1}; m_i)$ with $\eta_i = -1 + 2x_i + \delta(x_i - \psi)_+$, five sample sizes $n \in \{20, 50, 100, 200, 500\}$, two binomial indices $m_i \in \{1, 10\}$ and two covariate distributions in $[0, 1]$, $x_i = i/n$ and $x_i \sim \mathcal{B}e(1, 2)$.

Table 1: *Empirical Type I error rates (10,000 replicates) of the Davies (D) and proposed Score (S) approach in testing for a breakpoint in segmented regression with binomial outcomes.*

n	size	$x_i = i/n$		$x_i \sim \mathcal{B}e(1, 2)$	
		D	S	D	S
$m_i = 1$					
20	1	1.7	0.9	1.9	1.2
	5	7.6	5.9	8.0	6.3
	10	14.8	12.1	15.5	12.5
50	1	1.5	1.2	1.2	1.0
	5	6.4	5.8	6.1	5.1
	10	13.1	11.3	12.1	10.8
100	1	1.0	1.0	1.0	1.0
	5	5.9	5.1	5.5	5.3
	10	11.7	10.4	11.6	10.7
$m_i = 10$					
20	1	1.1	1.2	0.9	1.1
	5	5.0	5.5	4.5	5.3
	10	9.8	10.4	8.9	10.7
50	1	0.8	1.0	0.9	0.9
	5	4.5	4.9	4.9	4.8
	10	8.7	9.9	10.0	10.2
100	1	0.7	1.0	1.0	1.1
	5	4.4	4.7	4.9	5.0
	10	8.7	9.7	9.5	10.1

For simulations under the null we assume $\delta = 0$, while for the power studies we set $\delta = -4$ with ψ at the 0.50 and 0.75 quantiles of the covariate distribution. For Binomial regression with logit link it is clearly $W = \text{diag}\{m_i \hat{\mu}_{0i}(1 - \hat{\mu}_{0i})\}_{i=1, \dots, n}$ where $\hat{\mu}_{0i}$ are the fitted values under the null hypothesis.

Table 1 reports the empirical type I errors at the conventional nominal levels 1%, 5%, and 10%. For binary outcomes ($m_i = 1$) and small samples ($n = 20$), the Davies test exhibits empirical levels higher than the nominal ones, with non negligible differences even at relatively larger samples ($n = 50$); this makes the Davies test useless for hypothesis testing in small samples with binary outcomes. On the other hand, the Score test performs well by overestimating slightly the error rates only for $n = 20$. At $n = 100$ the empirical levels approach

to the nominal values for both the Score and the Davies tests. For binomial observations ($m_i = 10$), things go much better as expected: the empirical type I errors are equal (for Score) or less (for Davies) than the nominal levels at any sample size. Empirical rate errors at large samples ($n = 200$ and $n = 500$) are also consistent with the nominal levels in both binary and binomial cases, and thus results are omitted.

Table 2: *Empirical powers (5,000 replicates) of the Davies (D) and the Score (S) tests in testing for a breakpoint in segmented regression with binomial outcomes. The nominal level is 5%.*

ψ	n	$x_i = i/n$		$x_i \sim \mathcal{Be}(1, 2)$	
		D	S	D	S
$m_i = 1$					
$q_{0.50}$	20	13.8	11.0	11.5	6.9
	50	17.9	17.6	11.9	10.4
	100	23.8	27.6	15.0	16.0
	200	38.4	48.7	25.2	28.3
	500	77.4	85.6	51.2	58.3
$q_{0.75}$	20	11.9	9.3	10.9	8.7
	50	11.4	10.3	11.9	10.9
	100	12.9	13.1	16.0	17.9
	200	18.9	20.0	24.9	30.9
	500	40.0	42.2	55.8	66.0
$m_i = 10$					
$q_{0.50}$	20	39.7	48.5	24.5	26.7
	50	86.1	91.9	50.9	56.9
	100	98.8	99.6	83.1	86.8
	200	99.9	99.9	98.6	99.1
	500	100.0	100.0	100.0	100.0
$q_{0.75}$	20	21.6	23.6	22.8	27.3
	50	43.6	45.5	53.1	61.7
	100	72.7	71.3	84.4	89.7
	200	95.9	94.0	99.1	99.4
	500	100.0	100.0	100.0	100.0

Table 2 portrays the rejection rates under the alternative. When examining the powers for binary data ($m_i = 1$), it should be emphasized that at small samples ($n = 20$, and also $n = 50$ to minor extend), Davies overestimates the type I error probability (see corresponding entries in Table 1), and therefore comparisons with Score are somewhat unfair. At larger samples, Score performs somewhat better than Davies with lighter differences in the binomial scenario. Here notice that we have also reported the useless results for $n = 500$ for completeness of the Table.

Poisson log-linear regression

To assess the performance of the proposed approach in testing for a breakpoint in Poisson log linear models, we assume $y_i \sim \text{Pois}(\exp \eta_i)$ where $\eta_i = 2 - .3x_i + \delta(x_i - \psi)_+$. As for the binomial case, we use five sample sizes $n \in \{20, 50, 100, 200, 500\}$, and two covariate distributions in $[0, 1]$, $x_i = i/n$ and $x_i \sim \mathcal{Be}(1, 2)$. Of course we set $\delta = 0$ for simulations under the null hypothesis, while we use $\delta = 0.6$ with ψ at quantiles 0.50 and 0.75 for the assessing the power. The weight matrix is $W = \text{diag}\{\hat{\mu}_{0i}\}_{i=1, \dots, n}$ where $\hat{\mu}_{0i}$ are fitted values from the linear model.

Table 3 shows the sizes corresponding to the usual nominal levels. Both Davies and Score returns correct results with empirical values equal to or lower than the nominal levels even at small sample sizes; empirical error rates for $n = 200$ and $n = 500$ are omitted.

Table 3: *Empirical Type I error rates (10,000 replicates) of the Davies (D) and proposed Score (S) approach in testing for a breakpoint in segmented regression with Poisson outcomes.*

n	size	$x_i = i/n$		$x_i \sim \mathcal{Be}(1, 2)$	
		D	S	D	S
20	1	0.9	1.2	0.9	1.0
	5	4.5	5.3	4.7	5.1
	10	8.9	10.7	9.4	10.0
50	1	0.9	0.9	0.7	0.9
	5	4.3	5.0	4.0	5.0
	10	8.6	10.1	9.1	9.9
100	1	0.9	1.1	0.8	1.0
	5	4.2	5.0	4.2	4.7
	10	8.5	9.8	8.5	9.7

Table 4: *Empirical powers (5,000 replicates) of the Davies (D) and the Score (S) tests in testing for a breakpoint in segmented regression with Poisson outcomes. The nominal level is 5%*

ψ	n	$x_i = i/n$		$x_i \sim \mathcal{Be}(1, 2)$	
		D	S	D	S
$q_{0.50}$	20	5.3	7.3	4.8	5.7
	50	8.2	11.9	7.6	9.5
	100	13.2	19.3	9.6	13.1
	200	23.5	58.4	16.9	22.8
	500	33.1	69.4	39.0	49.1
$q_{0.75}$	20	4.6	5.9	5.6	7.2
	50	5.6	6.9	6.8	8.6
	100	7.0	8.5	9.1	12.8
	200	10.7	12.8	16.1	23.3
	500	24.0	27.0	35.9	47.4

Table 4 illustrates the powers of the two tests. The proposed Score test always reports higher power, regardless of sample size, covariate distribution, and location of ψ ; differences between S and D are particularly marked at large samples when $x_i = i/n$ and $\psi = q_{0.50}$.

In this supplementary note we have presented further simulation results concerning performance of the proposed Score approach with respect to the Davies test when interest lies in testing for existence of a breakpoint in Poisson and Binomial GLMs.

In summary, simulations carried for Poisson and Binomial GLMs, emphasize that in all other considered scenarios the Score test exhibits similar or higher power than the Davies test. Noticeably the Score test yields reliable results even with binary data and small samples where the Davies test overestimates the type I error rates.