METRON (2013) 71:21–32
DOI 10.1007/s40300-013-0004-1

Novel symmetry tests in regression models based on Gini mean difference

Hend Auda

Received: 28 June 2010 / Accepted: 31 January 2013 / Published online: 28 May 2013

Abstract This article proposes two new tests of symmetry based on the Gini mean difference. The symmetry hypothesis of the disturbance in a linear regression model around zero was analyzed using the proposed tests. A Monte Carlo simulation study shows that the tests have good size and power properties for sample sizes as small as 30. The symmetry of the error term in a cross county model of Gini index as a measure of income inequality and consumer price inflation was studied by the proposed tests.

Keywords Symmetry · Gini mean difference · Simulation · Cross country model

1 Introduction

Consider the following linear regression model:

$$Y = X\beta + \epsilon \tag{1}$$

where X is $(n \times d)$ regressor matrix that has 1's in the first column to represent the intercept, β is $(d \times 1)$ regression coefficients vector, Y is $(n \times 1)$ regress and vector and ϵ is $(n \times 1)$ error vector. Under the multiple linear regression assumptions, ϵ is assumed to be independently distributed and follows a normal distribution with mean zero and constant variance σ^2. Based on these assumptions (homoscedasticity, independent and normally distributed errors), the ordinary least squares (OLS) estimators for the regression coefficients are the best linear unbiased estimators (BLUE) for the regression parameters and they are identical to the maximum likelihood estimators.

When the error distribution is non-normal, the efficiency of the OLS estimators deteriorates. One approach to deal with non-normal error distributions is the adaptive estimation.

H. Auda (✉)
Statistics Department, Helwan University, Helwan, Egypt
e-mail: hendauda@yahoo.com

H. Auda
The Egyptian Cabinet of Ministers, Information and Decision Support Center, Cairo, Egypt

Bickel [2] has shown that the symmetry of the distribution of errors around zero is an important assumption in constructing the adaptive estimators when the error distribution is unknown. Newey [9] used a generalized method of moments approach for adaptive estimation based on the symmetry assumption for the error term.

Fan and Gencay [4] proposed a non-parametric test of the symmetry hypothesis for the error term in a linear regression model. Their test is based on the kernel estimation of the density function of the errors. Ahmad and Li [1] proposed a new test for symmetry of regression errors also based on kernel estimation. Although the latter tests are distribution free consistent tests, their smoothing methods may affect the size and power in finite samples [12]. The simulation in both studies showed that the tests have good properties for sample sizes as small as 50.

This article proposes two new distribution free tests of symmetry in the linear regression models. The new tests are based on the Gini mean differences and the properties of these tests are compared with the one proposed by Fan and Gencay [4]. There are many other tests of symmetry previously introduced in the literature. However, in each study, a comparison of few selected competitors is provided. Another comprehensive simulation study on the 14 most popular tests of symmetry, in which two new proposed tests of symmetry based on Gini mean difference are introduced, is in [13]. In this study, we compare type I error probabilities and power for 17 null and alternative distributions. The proposed tests (in particular a rank based test RS) compare favorably with several existing procedures in controlling the type I error as well as in power. In [13], an important novelty figures 1–3 in pages 10–12 show the comparison results of type I error probabilities and power of the 16 tests for 17 null and alternative distributions.

Gini [6] proposed the mean difference (GMD) as a measure of variability and Gini [7] introduced the Gini index (GI), as a measure of income inequality. The Gini mean difference is defined as

$$GMD = \frac{\sum_{i<j}^{n} |x_i - x_j|}{n(n-1)/2} \tag{2}$$

In Sect. 2, the proposed tests and their statistical inferences are defined. The Monte Carlo simulation study that was conducted to examine the performance of the proposed tests with an empirical example is given in Sect. 3.

2 The proposed tests of symmetry

The proposed tests of symmetry for the regression error term are based on a comparison of variability in the data below and above zero measured by the GMD (2). One of the proposed tests is applicable to the original data, and the other has its rank-based counterpart.

As mentioned above, we are interested in testing the symmetry of the regression error term around zero, such that $H_0 : f(e\,|x) = f(-e\,|x)$ against the alternative $H_1 : f(e|x) \neq f(-e|x)$.

For simplicity, we will assume that the regressors are fixed, and construct a test for $H_0 : f(e) = f(-e)$against the alternative $H_1 : f(e) \neq f(-e)$.

Let $e_1,\ldots, e_n$ be a standardized random sample from unknown continuous distribution. Partition the data into two subsets: subset1 ($e_i > 0$) and subset2 ($e_i < 0$), and since the population distribution is continuous, $P(e_i = 0) = 0$ and such values should occur very rarely in the data set. If they happen to appear, they can be excluded from the analysis, and the sample size would be reduced accordingly.

As a test statistic, we take

$$S = \frac{GMD_+}{GMD_-} \tag{3}$$

where GMD_+ and GMD_- are the Gini mean difference statistics given by formula (2) applied to subset 1 and 2, respectively.

A similar approach can be used for the ranks instead of the actual data values. Thus, we obtain

$$RGMD_+ = \frac{2\sum_{k=1}^{n_1} kR^1(k) - (n_1+1)\sum_{k=1}^{n_1} R^1(k)}{n_1(n_1-1)/2} \tag{4}$$

where n_1 is the sample size for subset 1 and $R^1(k)$ is the rank of the k-th in magnitude value in subset 1 when all $e_1, e_2, \ldots, e_n$ are ranked jointly. Similarly

$$RGMD_- = \frac{2\sum_{j=1}^{n_2} jR^2(j) - (n_2+1)\sum_{j=1}^{n_2} R^2(j)}{n_2(n_2-1)/2} \tag{5}$$

where n_2 is the sample size for subset 2 and similarly $R^2(j)$ is the rank of the k-th in magnitude value in subset 2 when all $e_1, e_2, \ldots, e_n$ are ranked jointly. As a test statistic we take

$$RS = \frac{RGMD_+}{RGMD_-} \tag{6}$$

2.1 Test based on the original data S

The null distribution of the test statistic that we propose for the original data depends on the type of the population distribution. For a large sample, the asymptotic distribution of the S statistic in (4) is normal but the parameters depend on the underling population distribution (see [11]). To show that, lets derive the parameters of the asymptotic distribution of the GMD in the following cases:

1. Derivation of parameters of the asymptotic distribution of the GMD in the case of the U[0, $\sqrt{12}$] and EXP(1) distributions

GMD could be expressed as L-estimate in the form:

$$T_n = \sum_{i=1}^{n} c_{ni}^* x_{i:n}, \tag{7}$$

where $c_{ni}^* = \frac{2(2i-n-1)}{n(n-1)}$.

Chernoff et al. [3] obtained the limiting distribution for the L-estimate in the general form:

$$T_n = \frac{1}{n}\sum_{i=1}^{n} c_{ni} h\left(x_{i:n}\right), \tag{8}$$

where h is some measurable function and $c_{ni} = \frac{2(2i-n-1)}{(n-1)}$ They showed that the asymptotic distribution for T_n is normal with mean μ_n in the form:

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} c_{ni} \tilde{H}\left(\tilde{v}_{ni}\right), \tag{9}$$

where F is the cdf of the distribution of x_i, $i = 1, \ldots, n$ and $\tilde{H}(\tilde{v}_{ni}) = F^{-1}(G(\tilde{v}_{ni}))$, $\tilde{v}_{ni} = \sum_{j=1}^{i} \left(\frac{1}{n-j+1}\right)$ and $G(x) = 1 - e^{-x}$, and variance σ_n^2 in the form:

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} \alpha_{ni}^2, \tag{10}$$

where $\alpha_{ni} = \frac{1}{n-i+1} \sum_{j=i}^{n} c_{nj} \tilde{H}'(\tilde{v}_{nj})$ and $\tilde{H}'$ is the derivative of the function $\tilde{H}$.

For $h(x) = x$ and $c_{ni} = \frac{2(2i-n-1)}{(n-1)}$ we could derive the asymptotic distribution for the GMD as a special case of L-estimates. We obtain means and variances of the GMD in the case of two different distributions : U[0, $\sqrt{12}$] and EXP(1).

A. $X_1, X_2, \ldots X_n \sim U[0, \sqrt{12}]$.

The asymptotic mean of the GMD:

$F(x) = \frac{x}{\sqrt{12}}$ and $F^{-1}(x) = \sqrt{12}x$. Using (3.10) and (3.11),

$$\tilde{H}(v_{nj}) = F^{-1}(G(\tilde{v}_{ni})) = F^{-1}\left(1 - e^{-\tilde{v}_{nj}}\right) = \sqrt{12}\left(1 - e^{-\tilde{v}_{nj}}\right) \tag{11}$$

Therefore, the asymptotic mean (3.10) can be derived as:

$$\mu_n = \frac{2\sqrt{12}}{n(n-1)} \sum_{j=1}^{n} (2j - n - 1)\left[1 - e^{-\tilde{v}_{nj}}\right] \tag{12}$$

$$\mu_n = \frac{2\sqrt{12}}{n(n-1)} \left[2\sum_{j=1}^{n} j - n^2 - n\right] - \frac{2\sqrt{12}}{n(n-1)} \sum_{j=1}^{n} (2j - n - 1)e^{-\tilde{v}_{nj}} \tag{13}$$

Since $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$, we obtain

$$\mu_n = -\frac{2\sqrt{12}}{n(n-1)} \sum_{j=1}^{n} (2j - n - 1)e^{-\tilde{v}_{nj}}. \tag{14}$$

The asymptotic variance of the GMD:

In the general case:

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} \alpha_{ni}^2, \tag{15}$$

where

$$\alpha_{ni} = \frac{1}{n-i+1} \sum_{j=i}^{n} c_{nj} \tilde{H}'(\tilde{v}_{nj}). \tag{16}$$

Since

$$\tilde{H}(\tilde{v}_{nj}) = \sqrt{12}\left(1 - e^{-\tilde{v}_{nj}}\right)$$

we have

$$\tilde{H}'(\tilde{v}_{nj}) = \sqrt{12}e^{-\tilde{v}_{nj}} \tag{17}$$

and finally

$$\alpha_{ni} = \frac{2\sqrt{12}}{n-i+1} \sum_{j=i}^{n} \frac{(2j-n-1)}{(n-1)} e^{-\tilde{v}_{nj}} \tag{18}$$

B. $X_1, X_2, \ldots, X_n \sim$ EXP(1).

The asymptotic mean of the GMD:

The cdf $F(x) = 1 - e^{-x}$, $F^{-1}(x) = -\log(1-x)$. The asymptotic distribution of GMD is normal. In the exponential case, $F^{-1}(x) = -\log(1-x)$ so that

$$\tilde{H}(v_{ni}) = F^{-1}\left(G\left(\tilde{v}_{ni}\right)\right) = -\log(1 - 1 + e^{-\tilde{v}_{ni}}) = \tilde{v}_{ni} \sum_{j=1}^{i} \frac{1}{n-j+1}, \tag{19}$$

and the asymptotic mean is derived as:

$$\mu_n = \frac{2}{n(n-1)} \sum_{i=1}^{n} (2i-n-1)\tilde{v_{ni}}. \tag{20}$$

The asymptotic variance of the GMD: The asymptotic variance is derived:

$$\alpha_{ni} = \frac{1}{n-i+1} \sum_{j=i}^{n} c_{nj} \tilde{H}'(\tilde{v}_{nj}) \tag{21}$$

We have

$$\tilde{H}(\tilde{v}_{nj}) = \tilde{v}_{nj}, \tag{22}$$

therefore

$$\tilde{H}'(\tilde{v}_{nj}) = 1 \tag{23}$$

and finally

$$\alpha_{ni} = \frac{1}{n-i+1} \sum_{j=i}^{n} \frac{2(2j-n-1)}{(n-1)}. \tag{24}$$

Since

$$\sum_{j=i}^{n} (2j-n-1) = 2\sum_{j=i}^{n} j - (n+1)(n-i+1) = (i+n)(n-i+1)$$

$$(n-i+1)(i+n-n-1) = (i-1)(n-i+1),$$

we have

$$\alpha_{ni} = \frac{2(n-i+1)(i-1)}{(n-1)(n-i+1)} = \frac{2(i-1)}{n-1}. \tag{25}$$

and

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{2(i-1)}{n-1}\right)^2. \tag{26}$$

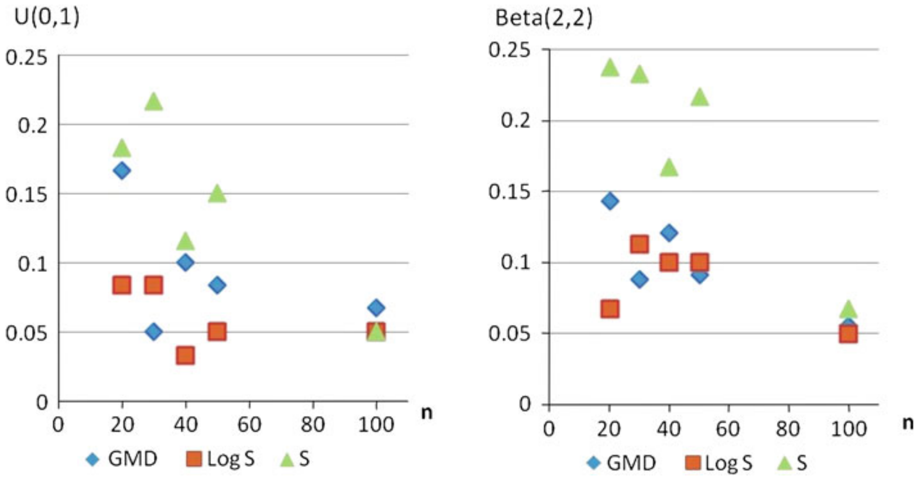

Fig. 1 The convergence of the distributions of GMD, Log S and S statistics to normality

When we compare the asymptotic mean and variance for the GMD in the uniform and exponential cases, we can see that the formulas are different which means that the parameters μ_n and σ_n^2 do depend on the underling distribution, similarly S and $\log(S)$ also do.

Beside depending on the underline distribution, convergences to normality for GMD and for the proposed statistic are slow (see Fig. 1).

Asymptotic normality of GMD, log(S) and S statistics

The rate of convergence of GMD depends on the underling distribution but in general it is rather slow. To study the convergence of the distributions of GMD, Log S and S statistics to normality, data from U[0,1] and Beta(2,2) were generated with different sample sizes, the proportion of cases when the normality assumption is rejected for the statistics based on 60 repetitions were obtained and graphically presented in Fig. 1.

For $n \leq 50$ the normality assumption is rejected. The normality can be assumed for $n \geq 50$ but it also depends on the distribution type (it is faster for uniform distribution, see Fig. 1). Thus, using asymptotic normality should not be recommended for $n < 50$ even if the parameters μ_n and σ_n are easy to obtain. Therefore, to test the symmetry based on original data we recommend using a permutation test or possibly a bootstrap test.

The test statistic that we propose for the original data will be permutation tests applied to (4). The permutation tests were first introduced as a non-parametric alternative to the two-sample t-test. Pitman estimator was based on the permutation test for the two-sample problem, block design and simple linear regression. Legendre and Anderson [8] showed that some permutation tests have equal or even more power than their parametric counterpart.

Permutation tests find a p-value the proportion of values of a test statistic at least as extreme as its original value (among values obtained by all possible partitions of the data with the same sample sizes n_1 and n_2). If the sample size n is too large, and instead of considering all possible permutations to estimate a p-value, permutations are randomly generated. Under the null hypothesis of symmetry distribution of S and $1/S$ are the same (if $n_1 = n_2$) or very close and one can estimate the p-value for the S statistic as:

$$\hat{P}(S \geq s) + \hat{P}(S \leq 1/s),$$

Table 1 RS statistic for all combinations for $n = 5$

n_1	n_2	Ranks in subset1	Ranks in subset2	rs
2	3	1, 2	3, 4, 5	0.750
		1, 3	2, 4, 5	1.000
		1, 4	2, 3, 5	1.500
		1, 5	2, 3, 4	3.000
		2, 3	1, 4, 5	0.375
		2, 4	1, 3, 5	0.750
		2, 5	1, 3, 4	1.500
		3, 4	1, 2, 5	0.375
		3, 5	1, 2, 4	1.000
		4, 5	1, 2, 3	0.750
2	2	1, 2	4, 5	1.000
		1, 4	2, 5	1.000
		1, 5	2, 4	2.000
		2, 4	1, 5	0.500
		2, 5	1, 4	1.000
		4, 5	1, 2	1.000

or as

$$2(\min(\hat{p}(S \geq s), \hat{p}(S \leq 1/s)))$$

where $\hat{P}$ are relative frequencies from the randomly generated permutations.

2.2 Inference for rank-based statistic RS

Inference for the test based on ranks may be based on the exact distribution. First, GMD is undefined if all the data are below or above the symmetry point and this is most likely when n is small. In addition, since one needs at least two observations in order to obtain the GMD, we considered $n = 5$ as the smallest sample size. Sample size $n = 5$ will be used as an illustrative example for interpreting how the exact distribution was obtained.

We obtain the exact distribution for the rank-based statistic in a way similar to the following example: we first list all possible combinations (n_1, n_2) for the sample sizes of subsets 1 and 2. In the case of $n = 5$, we have two cases: either $n_1 = 2$, $n_2 = 3$ or $n_1 = 3$, $n_2 = 2$. We then compute the statistic RS as given in Eq. (2). All possible combinations for $n = 5$ are provided in Table 1 with the corresponding values for the RS statistic.

The exact distribution for the statistic when $n = 5$ is given in Table 2:

For testing symmetry, we use a two sided test as we are testing whether the distribution is symmetric against the general alternative. To compute the p-value according to the above table, we have

$$P\text{-value} = P(RS \geq rs) + P(RS \leq 1/rs)$$

As for $P(RS \leq 1/rs)$, we can compute it directly from the table. However, for $P(RS \geq rs)$, we can compute it as $1 - P(RS < rs)$. For $P(RS < rs)$, we can use the closest value for the statistic that is less than rs. The R-code used to generate the critical values is provided in the Appendix.

Table 2 Cdf for the null distribution of the RS statistic

n_1	n_2	rs	$P(RS = rs)$	$P_0(RS \leq rs)$
2	3	0.375	0.2	0.2
		0.75	0.3	0.5
		1	0.2	0.7
		1.5	0.2	0.9
		3	0.1	1
2	2	0.5	0.167	0.167
		1	0.667	0.833
		2	0.167	1

3 Simulation study and empirical example

In this section, the size and power of the proposed tests of symmetry are investigated. Simulation is conducted in a way similar to [4], in order to compare the results of the proposed tests with theirs. The proposed tests are also used to test the symmetry of the error term in a model with cross-country data on inflation and income inequality measured by the Gini index.

3.1 Simulation study

In the simulation study, 1000 replications were used for each sample size and error distribution. Calculation of the proposed test of symmetry based on the original data S (as mentioned in Sect. 2) is done using the permutation technique. As for the other proposed test of symmetry based on ranks RS, the critical values were computed for each sample size.

Table 3 presents the simulation results with two regressors (including an intercept term).The tests S and RS have the desired empirical size for $\alpha = 0.05$ and for $\alpha = 0.1$,

Table 3 Size and power of the tests; (S) and (RS) with two regressors

n	20		30		50		100	
α	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
S								
t_5	0.067	0.133	0.068	0.106	0.058	0.109	0.068	0.11
$N(0, 1)$	0.055	0.093	0.049	0.094	0.053	0.110	0.048	0.109
Laplace	0.048	0.094	0.057	0.118	0.060	0.102	0.065	0.120
χ^2_2	0.498	0.647	0.873	0.933	0.991	0.997	1	1
Lognormal	0.617	0.729	0.950	0.982	1	1	0.999	1
Exponential	0.560	0.738	0.880	0.934	0.988	0.994	1	1
RS								
t_5	0.066	0.133	0.053	0.104	0.053	0.108	0.060	0.106
$N(0, 1)$	0.042	0.087	0.059	0.108	0.057	0.107	0.059	0.101
Laplace	0.048	0.121	0.061	0.102	0.060	0.120	0.061	0.103
χ^2_2	0.195	0.302	0.466	0.595	0.7867	0.831	0.974	0.981
Lognormal	0.240	0.417	0.545	0.651	0.824	0.867	0.981	0.988
Exponential	0.280	0.422	0.495	0.621	0.762	0.835	0.973	0.980

Table 4 Size and power of the tests (S) and (RS) with four regressors

n	20		30		50		100	
α	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
S								
t_5	0.059	0.104	0.064	0.102	0.067	0.121	0.061	0.132
$N(0, 1)$	0.055	0.105	0.038	0.088	0.045	0.094	0.045	0.095
Laplace	0.043	0.102	0.062	0.114	0.061	0.121	0.064	0.128
χ_2^2	0.341	0.581	0.688	0.806	0.974	0.989	1	1
Lognormal	0.430	0.603	0.809	0.893	0.988	0.993	1	1
Exponential	0.312	0.561	0.700	0.811	0.972	0.985	1	1
RS								
t_5	0.050	0.094	0.055	0.110	0.046	0.090	0.051	0.100
$N(0, 1)$	0.047	0.097	0.046	0.102	0.043	0.112	0.040	0.091
Laplace	0.046	0.119	0.058	0.124	0.052	0.114	0.055	0.101
χ_2^2	0.154	0.274	0.285	0.409	0.306	0.473	0.950	0.973
Lognormal	0.199	0.331	0.311	0.436	0.663	0.801	0.963	0.980
Exponential	0.178	0.299	0.297	0.404	0.580	0.728	0.946	0.970

Table 5 Size and power of the tests (S) and (RS) with six regressors

n	20		30		50		100	
α	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
S								
t_5	0.055	0.116	0.063	0.114	0.061	0.129	0.059	0.1030
$N(0, 1)$	0.054	0.107	0.051	0.104	0.041	0.095	0.047	0.097
Laplace	0.053	0.100	0.058	0.118	0.064	0.123	0.051	0.101
χ_2^2	0.175	0.324	0.540	0.698	0.928	0.972	1	1
Lognormal	0.177	0.307	0.625	0.760	0.965	0.988	1	1
Exponential	0.166	0.302	0.578	0.732	0.940	0.972	1	1
RS								
t_5	0.061	0.131	0.049	0.086	0.056	0.103	0.049	0.110
$N(0, 1)$	0.044	0.101	0.064	0.092	0.060	0.113	0.051	0.099
Laplace	0.043	0.096	0.051	0.094	0.062	0.121	0.050	0.112
χ_2^2	0.137	0.199	0.268	0.352	0.559	0.680	0.941	0.967
Lognormal	0.144	0.207	0.301	0.369	0.590	0.728	0.952	0.974
Exponential	0.132	0.212	0.256	0.338	0.534	0.672	0.933	0.962

respectively. As for sample sizes $n = 50$ and 100, the tests have good power for all asymmetric alternatives considered in the simulation. In case of a small sample size ($n = 30$), the S test has good power and its power is higher than the RS test.

Tables 4 and 5 extend the size and power calculations of the tests with the number of regressors $d = 4$ and $d = 6$. The number of regressors is increased to study its effect on the

Table 6 The OLS estimates of the regression model

Variable	Coefficient	Standard error	t-ratio	P-value
Constant	43.95	1.778	24.717	$<2e-16$
Δcpi	−0.0899	0.033	−2.724	0.009
Δcpi^2	0.0002	6.126e−05	2.887	0.006

power of the proposed tests. For sample size 100, the increase in the number of regressors does not alter the power of the S test. For sample sizes 50, 30 and 20, the largest reduction in power of the S test is 6, 20 and 47 %, respectively. The S test is rather stable to an increase in the number of regressors for sample size 30 and above. For RS test with sample sizes 100 and 50, the largest reduction in power of the RS test is 23 %. For sample sizes 30 and 20, the largest reduction in power of the RS test is 25 %. The RS test although it has less power than S test for small sizes (20 and 30); it is stable to an increase in the number of regressors for sample size 20 and above. The percentage of overrejection or underrejection of the empirical size is within 2 % of its nominal size at both 5 and 10 % levels.

For sample sizes 100 and 50, the proposed tests of symmetry have similar high power as the test proposed by Fan and Gencay (1995). The special property for the S statistic is that it has good power for size less than 50 ($n = 30$), while Fan and Gencay test cannot be applied for sizes less than 50.

3.2 An empirical example

The hypothesized existence of a non-monotonic relationship between inequality and inflation [5] is examined by estimating a quadratic model of the form:

$$\Delta GI = \alpha + \beta_1 \Delta cpi + \beta_2 \Delta cpi^2 + \epsilon \tag{27}$$

where Δ GI is the annual average of Gini index, Δ cpi is the average annual of consumer price inflation across 47 developing countries, gathered from World Development Indicators (2008), CD-ROM, World Bank, Washington D.C. during the period 1984–2005. All observations are annual averages for the sample period.

In Table 6 the regression estimates are statistically significant. The F-statistic which is equal to 4.229 with p-value = 0.0209 means that the overall estimated equation is significant at 5 % level of significance.

To test the normality of the residuals, the Shapiro normality [10] test is equal to 0.9492 with p-value = 0.04, which means that the normality assumption is rejected at 5 % level of significance. The S test is equal to 1.219 with p-value = 0.178, which means that the symmetry assumption is retained at 5 % level of significance. The RS test is equal to 0.834 and the null hypothesis of symmetry is retained at 10 % level of significance. As the symmetry assumption for the errors is not rejected, the regression coefficients can be estimated adaptively by the generalized method of moments as that of Newey [9] to get maximum efficient estimators for the model.

4 Conclusion

Two new tests of symmetry based on the Gini mean difference were used to identify the symmetry of the residual of the linear regression model. The proposed test of symmetry S converges slowly to the normal distribution with parameters depending on the underlying population. The permutation tests based on statistic S were proposed, and proved good size and power properties in sample sizes as small as 30 observations. The other proposed test, RS, is the rank-based counterpart and its exact distribution was determined. The test has good size and power properties in samples of size not less than 50 observations.

The relationship between the Gini index and the Consumer Price Inflation for cross county data was investigated, and the symmetry of the errors was tested with the proposed tests.

Appendix

R-codes

Below are some codes that were used in simulations, the codes compute RS statistic and its critical values.The follwoing are the main R-functions used in the code:

x–numeric vector of non-missing data points
median. – value of median (if known), the default is to estimate it with the sample median
abs(x)—absolute value of x
xpos—numeric vector of the positive values of x
xneg—numeric vector of th negative values of x
RGMD2—-Gini mean differences based on ranks
Returned value–RS statistic as ratio of GMD for positive and negative values
RScdf——cummulative function for the RS statistic
RSpf—-probability function for the RS statistic

1. RS statistic:

```
‘RS‘ < − function(x, median.=median(x),equal=F,ties.method=“first”)
{RGMD2 < − function(y, n)
sum((2*(1:n)-n-1) * y) / n / (n-1)
x < − x - median.
rx < − rank(abs(x),ties.method=ties.method)
xpos < − if(equal) x>=0 else x>0; n1 < − sum(xpos)
xneg < − x<=0; n2 < − sum(xneg)
rpos < − sort(rx[xpos])
rneg < − sort(rx[xneg])
RGMD2(rneg, n2) / RGMD2(rpos, n1)}
```

2. RS cdf:

```
‘RScdf‘ < −
function(N, n=(N/2), remove.median=(N>(2*n)))
{rs < − RSpmf(N,n,remove.median)
x < − as.numeric(names(rs))
y < − c(0, cumsum(rs))
pars < − c(N=N, n1=n, n2=N-n-remove.median)
result < − list(parameters=pars, x=x, pmf=rs[],
fun=stepfun(x, y, right=FALSE))
class(result) < − “RS” result}
```

3. RS pf:

```
'RSpmf' < −
function(N, n=(N/2), remove.median=(N>(2*n)))
{ RGMD2 < − function(y, n, ind=2*(1:n))
sum((ind-n-1) * y) / n / (n-1)
r < − 1 : N
if (remove.median) {
r < − r[-(n+1)]
N < − N - 1}
n2 < − N - n
rs < − numeric( ncombs < − choose(N,n) )
s1 < − 2 ∗ (set1 < − 1:n); s2 < − 2*(1:n2)
for(i in 1:ncombs){
rs[i] < − RGMD2(r[set1], n, s1) / RGMD2(r[-set1], n2, s2)
set1 < − nextcombo(set1, N, n)x}
table(rs) / ncombs }
```

References

1. Ahmad, I.A., Li, Q.: Testing symmetry of an unknown density function by kernel Method. J. Nonparametric Stat. **7**, 279–293 (1997)
2. Bickel, P.J.: On adaptive estimation. Ann. Stat. **10**, 647–671 (1982)
3. Chernoff, H., Gastwirth, L.J., Johns, V.M.: Asymptotic distribution of linear combination of functions of order statistics with applications of estimation. Ann. Math. Stat **38**(1), 52–72 (1967)
4. Fan, Y.: A consistent nonparametric test of symmetry in linear regression models. J. Am. Stat. Assoc. **90**, 551–557 (1995)
5. Galli, R., Hoeven, R.: Is inflation bad for income inequality? The importance of the initial rate of inflation. Employ Paper **29**, 1–38 (2001)
6. Gini, C.: Variabilitá e mutabilitá. In: Pizetti, E., Salvemini, T. (eds.) Reprinted in memorie di metodologica statistica. Libreria Eredi Virgilio Veschi, Rome (1955) (1912)
7. Gini, C.: Sulla misura della concentrazione e della variabilitá dei caratteri. Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti, tomo LXXIII, pp. 1203–1248. English translation in, Metron, 2005, 63(1)
8. Legendre, P., Anderson, M.J.: Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol. Monogr. **69**, 1–24 (1999)
9. Newey, W.K.: Adaptive estimation of regression models via moments restrictions. J. Econ. **38**, 301–339 (1988)
10. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality complete samples. Biometrika **52**, 591–611 (1965)
11. Serfling, J.R.: Approximation Theorems of Mathematical Statistics. Wiley, New York (1980)
12. Stengos, T., Wu, X.: Information-Theoretic Distribution Tests with Application to Symmetry and Normality, pp. 1–22. Social science research network, Working Paper Series (2004)
13. Wang, J.C., Auda, H., Niewiadomska-Bugaj, M.: Comparing a new gini test with other symmetry tests when median is unknown. Commun. Stat. Simul. Comput. **42**, 1616–1627 (2013)