AN OMNIBUS TEST FOR HETEROSCEDASTICITY USING RADIAL STATIONARITY AND DATA DEPTH

Thomas J. Farrar
Cape Peninsula University of Technology
e-mail: *farrart@cput.ac.za*

The classical linear regression model is a very well-known and widely used statistical method. One of the assumptions on which the model's validity rests is that of constant error variance (homoscedasticity). Thus, heteroscedasticity testing plays an important role in linear regression model diagnostics. This study proposes an omnibus test for heteroscedasticity in the classical linear regression model using the notion of radial stationarity about the centre of the explanatory variable space, combined with the notion of data depth. The test procedure consists of constructing a spatially ordered series of residuals (after removing the deepest observations) that is then tested for weak stationarity. Monte Carlo simulations show that, when the Priestley-Subba Rao method is used as the stationarity test, the resulting 'radial stationarity' test outperforms the Breusch-Pagan Test and White's Test in terms of average excess power over size under a variety of heteroscedastic alternatives, in some cases by a wide margin. The size of the proposed test is not robust under non-normality, however, and two nonparametric stationarity tests performed poorly in the simulations.

Key words: Data depth, Heteroscedasticity, Linear regression, Monte Carlo simulation, Power analysis, Weak stationarity.

1. Introduction

The classical linear regression model is a very widely used statistical method for analysing relationships between a continuous dependent variable and one or more independent variables. The model can be specified in matrix form as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{Y}$ is a vector of n random dependent variables, $\boldsymbol{\beta}$ is a vector of p unknown parameters, $\boldsymbol{X}$ is a $n \times p$ matrix - conventionally with $\mathbf{1}_n$ as its first column and the other $k = p - 1$ columns containing vectors of fixed, observed explanatory variables $\boldsymbol{X}_j$, $j = 1, 2, \ldots, k$ - and $\boldsymbol{\epsilon}$ is a vector of random errors or disturbances. Under the Gauss-Markov assumptions, the ordinary least squares (OLS) method provides a best linear unbiased estimator (BLUE) of the parameter vector. One of the Gauss-Markov assumptions is that the errors are homoscedastic (have a constant variance), i.e. $\operatorname{Var}(\epsilon_i) = \sigma^2$ for all $i = 1, 2, \ldots, n$. The violation of this assumption—namely, heteroscedasticity—renders the standard errors of the OLS estimators biased and inconsistent (Wooldridge, 2016), thereby invalidating interval estimates and inferences based on the model. Heteroscedasticity testing thus plays an important role in model adequacy diagnostics in any application of linear regression. Numerous methods for detecting heteroscedasticity have been proposed over the past several decades; among the more well-known are White's Test, the Breusch-Pagan Test, and the Goldfeld-Quandt Test.

The aim of this study is to propose a test for heteroscedasticity that involves notions of stationarity and data depth. The tests proposed are 'omnibus' tests in that they simultaneously test for heteroscedasticity linked to any of the explanatory variables $\boldsymbol{X}_j$. The advantage of omnibus tests is that they require little information about the suspected form of the heteroscedasticity; the disadvantage is that they also do not provide information about the form of heteroscedasticity (if present), and thus do not suggest a remedial strategy.

In Section 2, we discuss common models for heteroscedastic errors and briefly describe some well-known heteroscedasticity tests. Section 3 motivates for and describes the new radial stationarity test. Section 4 compares the power of this test (in parametric and nonparametric versions) to White's Test and the Breusch-Pagan Test. Section 5 draws conclusions and highlights areas for future research.

2. Background Concepts

2.1 Models of Heteroscedasticity

Models of heteroscedasticity commonly assume that the error variance is related to one or more of the explanatory variables $\boldsymbol{X}_j$. Accordingly, we can define $\mathrm{Var}(\epsilon_i) = \sigma^2 h(X_{i1}, X_{i2}, \ldots, X_{ik})$ where $h(\cdot)$ is some real-valued function. Two important forms of $h(\cdot)$ are known as the additive model and the multiplicative model respectively (Griffiths and Surekha, 1986), and can be expressed thus:

$$\mathrm{Var}(\epsilon_i) = \sigma^2 (1 + \lambda_1 X_{i1} + \lambda_2 X_{i2} + \cdots + \lambda_k X_{ik})^2 = \sigma^2 \left(1 + \sum_{j=1}^{k} \lambda_j X_{ij}\right)^2, \tag{1}$$

$$\mathrm{Var}(\epsilon_i) = \sigma^2 X_{i1}^{\lambda_1} X_{i2}^{\lambda_2} \cdots X_{ik}^{\lambda_k} = \sigma^2 \prod_{j=1}^{k} X_{ij}^{\lambda_j}. \tag{2}$$

In either case the λ_j are real-valued constants, and a test of the null hypothesis of homoscedasticity H_0, which can in general be written as $\mathrm{Var}(\epsilon_i) = \sigma^2$, $i = 1, 2, \ldots, n$, is equivalent to $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$.

2.2 Well-Known Heteroscedasticity Tests

Among the well-known tests for heteroscedasticity that appear in econometrics textbooks (such as Gujarati and Porter, 2009) are those of Goldfeld and Quandt (1965), Breusch and Pagan (1979), and White (1980). Goldfeld and Quandt's method is not an omnibus test, since it requires prior knowledge of *one* $\boldsymbol{X}_j$ to which heteroscedasticity is related. Their procedure consists of ordering the observations in nondecreasing order of this 'potential deflator', removing c central observations, and applying OLS separately to the 'upper' and 'lower' subsets of the data. An F test is then used to compare the error variances of these two models; if the null hypothesis of equal error variances is rejected, we conclude that heteroscedasticity is present. Breusch and Pagan proposed a Lagrange multiplier test (BP herein) that involves fitting an auxiliary regression model using OLS with the squared residuals of the original model e_i^2 as the dependent variable and some exogenous explanatory variables $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_q$. White separately proposed a test (W herein) that is a special case of BP, in which $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_q$ are $\boldsymbol{X}_j$ and $\boldsymbol{X}_j^2$ for $j = 1, 2, \ldots, k$ and, optionally, the $\binom{k}{2}$ two-way interaction terms. In both tests, the test statistic is $T = nr_{\mathrm{aux}}^2$, where r_{aux}^2 is the multiple coefficient of determination from the auxiliary regression. Under the null hypothesis, $T \sim \chi^2(q)$ and we reject H_0 if $T > \chi^2_{\alpha,q}$; this would imply that the variation in squared residuals is well-explained by the explanatory variables.

2.3 Depth Functions

Data depth is an important concept in multivariate statistical analysis that refers to 'the "depth" or "outlyingness" of a given multivariate sample with respect to its underlying distribution' (Liu, Parelius and Singh, 1999, p. 783). It leads to the notion of a centre-outward ordering of points in multidimensional space (Zuo and Serfling, 2000). A depth function is any real-valued function that quantifies data depth, with 'inner' (deeper, more central) points being mapped to large values and 'outer' (shallower, more outlying) points being mapped to small values. Zuo and Serfling (2000) describe four desirable properties that a depth function should possess and assess how well certain commonly-used depth functions perform against these criteria. They find that the halfspace and projection depth functions 'appear to represent very favorable choices', while the Mahalanobis depth function is also competitive. The Mahalanobis depth function was chosen for this study because it is computationally inexpensive and almost never yields ties when working with continuous data.

The Mahalanobis depth function is based on the distance metric introduced by Mahalanobis (1936). Following Liu et al. (1999), we can define the sample version of the Mahalanobis depth function of a point x with respect to its empirical distribution thus:

$$MHD(\boldsymbol{x}) = \frac{1}{1 + (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \boldsymbol{S}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}})},$$

where $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$ are, respectively, the sample mean vector and sample covariance matrix. In this study, we are interested in the Mahalanobis depth of the $\boldsymbol{X}$ vectors of observed explanatory variable values relative to their empirical distributions.

2.4 Weak Stationarity of Stochastic Series

If we conceive of some way of ordering our residuals spatially, concepts from time series analysis can be applied by analogy to the resulting 'spatial series'. Of special interest here is the notion of 'weak stationarity' (defined for time series in Brockwell and Davis, 2002). If $\{e_s : s = 1, 2, \ldots, n\}$ represents a spatially ordered series of residuals with $\mathrm{E}\left(e_s^2\right) < \infty$, then we define the mean function and covariance function of $\{e_s\}$ respectively as $\mu_e(s) = \mathrm{E}(e_s)$ and $\gamma_e(r, s) = \mathrm{Cov}(e_r, e_s)$ for all integers r and s. The spatial series $\{e_s\}$ is said to be weakly stationary if $\mu_e(s)$ is independent of s and $\gamma_e(s + h, s)$ is independent of s for each h. The second condition with $h = 0$ is equivalent to the homoscedasticity assumption.[1]

3. The Radial Stationarity Test

If heteroscedasticity is present in the classical linear regression model in either the additive form in (1) or multiplicative form in (2), the magnitude of the residuals will be related to the empirical,

[1] Under the Gauss-Markov assumptions, the linear regression residuals e_i have mean 0 and variance $\sigma^2 m_{ii}$ where m_{ii} is the ith diagonal element of the matrix $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top$. Thus, for weak stationarity to hold the residuals e_i should be transformed to $\tilde{e}_i = e_i / \sqrt{m_{ii}}$.

spatial distribution of the $\boldsymbol{X}$ observations in two ways that can be observed in Figure 1 (where $k = 2$). First, since $h(\cdot)$ is monotonic in X_1 and X_2, the observations with outlying (low-depth) $\boldsymbol{X}$ values will tend to have extreme (large or small) error variances (note the small points in the bottom left region of the plot and large points in the top right region), while the observations with central (high-depth) $\boldsymbol{X}$ values will tend to have medium-sized error variances; in other words there will be less variation in the squared residuals the deeper in $\boldsymbol{X}$-space we go.[2] Secondly, if we construct a spatial series by moving about the centre of the $\boldsymbol{X}$-space through the outer points (i.e., below some depth threshold) in a circular fashion, it appears that the resulting series is not weakly stationary: $\gamma_{\tilde{e}}(s, s)$ depends on location in the $\boldsymbol{X}$-space and therefore on s. These two patterns give rise to a heteroscedasticity test that we will call the radial stationarity (RS) test.

An intuitive method of obtaining a 'radially ordered' series of residuals, in the case $k = 2$, is to obtain standardised $\boldsymbol{X}$ values $\boldsymbol{X}^\star = \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right) \boldsymbol{S}^{-1}$ and to determine the angle between each $\boldsymbol{x}_i^\star$ point and the horizontal line $X_1^\star = 0$, i.e., $\theta_{i,1,2} = \arctan\left(x_{i2}^\star / x_{i1}^\star\right)$ (as seen in Figure 1). The null hypothesis of weak stationarity of the resulting series is practically equivalent to the null hypothesis of homoscedasticity.[3] Because the residuals seem nearer to stationarity in the most central region of the $\boldsymbol{X}$-space, the power of this test to detect heteroscedasticity may be enhanced by omitting the deepest c points (in terms of MHD) from the series. These insights motivate the following hypothesis testing procedure for $k = 2$ independent variables:

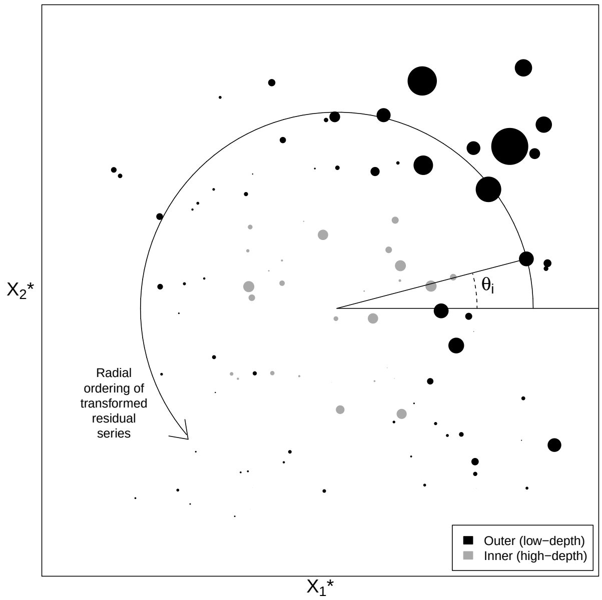

Figure 1. $\tilde{e}_i^2$ (proportional to point size) by standardised X_1 and X_2 values for a linear regression model with $k = 2$, $n = 100$, and $\mathrm{Var}(\epsilon_i) = X_{i1}^3 X_{i2}^3$.

[2] This motivates the possibility of a heteroscedasticity test that compares the residual *kurtosis* between low-depth and high-depth observations. However, sampling distributions of kurtosis statistics tend to be extremely sensitive to departures from normality, and since most of the contribution to sample kurtosis comes from a few extreme values, rank-based nonparametric methods for kurtosis comparison are unlikely to be effective.

[3] As a reviewer observed, if $\gamma_e(s + h, s)$ were independent of s for $h = 0$ but dependent on s for some $h > 0$ then homoscedasticity would hold but weak stationarity would not, due to autocorrelation. Thus the two hypotheses are not exactly equivalent, but since autocorrelation is another violation of linear regression model assumptions, this fact broadens the utility of the RS test as an omnibus diagnostic tool. The kind of autocorrelation in question is not serial autocorrelation, since the observations have been radially reordered. If $\boldsymbol{X}$ included spatial variables such as GPS coordinates, the RS test could have some power to detect spatial autocorrelation. Therefore, under the assumption of no 'radial autocorrelation' or in applications in which 'radial autocorrelation' would have no practical interpretation, the RS test could be treated primarily as a test of heteroscedasticity.

1. Fit OLS to the data and obtain transformed residuals $\tilde{e}_i = e_i/\sqrt{m_{ii}}$
2. Standardise the $\boldsymbol{X}$ observation matrix and remove the c deepest points (i.e. with largest MHD).
3. Create a series of transformed residuals ordered by increasing $\theta_{i,1,2}$.
4. Conduct a test of weak stationarity on the series; if the null hypothesis of weak stationarity is rejected then the null hypothesis of homoscedasticity is also rejected.

If $k > 2$ it is problematic to construct a single 'radially ordered' series of residuals in standardised $\boldsymbol{X}$-space. However, noting that the $\boldsymbol{X}$-space consists of $\binom{k}{2}$ Cartesian planes, one can cycle through these planes and construct $\binom{k}{2}$ series of residuals that are radially ordered by angle to the respective planes. There are then $\binom{k}{2}$ weak stationarity tests to conduct, and one then needs to perform a Bonferroni or other correction to control the family-wise Type I error rate (the overall size of the test). It is, however, redundant to cycle through *all* the $\binom{k}{2}$ planes. If one considers first the X_1X_2 plane, then the X_3X_4 plan, and so on up to $X_{k-1}X_k$ plane, then all X_j will have been included in the test. It is thus sufficient to cycle through $\pi = k/2$ planes if k is even, and $\pi = (k-1)/2 + 1$ planes if k is odd (in which case one X_j must be included twice). The test procedure for $k > 2$ is:

1. Fit OLS to the data and obtain transformed residuals $\tilde{e}_i = e_i/\sqrt{m_{ii}}$
2. Standardise the $\boldsymbol{X}$ observation matrix and remove the c deepest points (i.e. with largest MHD).
3. Create a series of transformed residuals ordered by increasing $\theta_{i,1,2}$.
4. Conduct a test of weak stationarity on the series.
5. Repeat (3) and (4) for the X_3X_4 plane and so on until the πth plane (where all X_j have been included at least once).
6. If the smallest P-value $< \frac{\alpha}{\pi}$ (where α is the desired test size), reject the null hypothesis of homoscedasticity.

Two stationarity testing methods were considered: the spectral density method of Priestley and Subba Rao (1969) (which is parametric in that it assumes normality of the series), abbreviated RS(PSR), and the recent nonparametric test of Jin, Wang and Wang (2015), abbreviated RS(JWW).

4. Monte Carlo Simulations

Monte Carlo simulations were used to estimate the power and size of the radial stationarity test with each of three stationarity test methods.[4] For comparison purposes, the Breusch-Pagan test (using the $\boldsymbol{X}_j$ as our exogenous explanatory variables) and White's Test (without cross-terms) were also simulated. The χ^2 degrees of freedom for these versions of BP and W are k and $2k$ respectively. BP has been used as a benchmark in previous simulation studies on the power of heteroscedasticity tests (Griffiths and Surekha, 1986; Račkauskas and Zuokas, 2007).

[4] The R package `fractal` of Constantine and Percival (2017) was used for RS(PSR). R code for RS(JWW) is referenced in Jin et al. (2015).

Simulations were run in R statistical software with 10^4 replicates for each case. Since none of the tests consistently achieved the nominal test size even with normally distributed errors (see Table 3), comparing the estimated power of all the tests at some fixed nominal size makes little sense. Rather, following a reviewer suggestion, the average height of the receiver operating-characteristic (ROC) curve of test t, $\widehat{W}_t(l,u)$, was estimated over an interval of relevant sizes using the method of Lloyd (2005). The average size in this interval, $(l+u)/2$, was subtracted from $\widehat{W}_t(l,u)$ to obtain $\widehat{Q}_t(l,u)$, the average excess power over size. In this case we used $l = 0.01$ and $u = 0.1$. The advantages of this method are that (i) it adjusts for departures from nominal test size; and (ii) it averages the tests' performance across a range of relevant sizes rather than only at one particular size. Since the $\widehat{Q}_t$ were all estimated using the same simulated data, their differences are dependent and therefore, again following Lloyd (2005), the standard errors of these differences were estimated using nonparametric bootstrap sampling with 100 replications.[5]

The X_j observations were generated from the uniform$(1,3)$ distribution independently, with $k = 2$, 4, and 8. Errors were generated from the normal distribution as well as the Laplace and uniform distributions in order to assess the robustness of the tests against leptokurtic and platykurtic departures from the normality assumption. Only the sample size $n = 100$ was used due to the number of other parameters to be varied. Various heteroscedastic models were attempted by changing the λ_j values in (1) and (2). In Table 2, '1A' and '1M' refer to the additive and multiplicative heteroscedasticity models respectively with λ_j values as per Case 1 in Table 1.[6] RS(PSR) estimates shown in Table 2 represent $c = 0.3n$ points removed for all cases; RS(JWW) estimates represent $c = 0$ points removed for all cases. The same conditions were used for size estimation simulations shown in Table 3.

Table 1. Different heteroscedastic models simulated.

Case No.	λ Coefficients	Description
0	All $\lambda_j = 0$	Homoscedasticity (Null Hypothesis)
1	One $\lambda_j = 1$ (all others 0)	Heteroscedasticity linked to exactly one X_j
2	All $\lambda_j = 1$	Heteroscedasticity linked to all X_j with same direction and magnitude
3	All $\lambda_j = \pm 1$ (half negative, half positive)	Heteroscedasticity linked to all X_j with same magnitude but in different directions
4	All $\lambda_j > 0$ (half = 1, half = $\frac{1}{2}$)	Heteroscedasticity linked to all X_j with same direction but different magnitude
5	All $\lambda_j \neq 0$ (half = ± 1, half = $\pm\frac{1}{2}$, equally divided between positive and negative)	Heteroscedasticity linked to all X_j with different directions and different magnitudes
6	Half of $\lambda_j = 1$ (all others 0)	Heteroscedasticity linked to half of X_j with same direction and magnitude
7	Half of $\lambda_j = \pm 1$ (half of these negative, half positive)	Heteroscedasticity linked to half of X_j with same magnitude but in different directions
8	Half of $\lambda_j > 0$ (half of these = 1, half = $\frac{1}{2}$)	Heteroscedasticity linked to half of X_j with same direction but different magnitudes
9	Half of $\lambda_j \neq 0$ (half = ± 1, half = $\pm\frac{1}{2}$, equally divided between positive and negative)	Heteroscedasticity linked to all X_j with different directions and different magnitudes

4.1 Simulation Results

Monte Carlo simulation estimates of average excess power over size ($\widehat{Q}$) in the size interval $[0.01, 0.1]$ are shown in Table 2, along with bootstrap estimates of the standard errors of $\widehat{Q}_{\mathrm{RS(PSR)}} - \widehat{Q}_{\mathrm{BP}}$. Monte Carlo simulation results of test size for all three error distributions are displayed in Table 3.

[5]These standard error estimates are reported only for $\widehat{Q}_{\mathrm{RS(PSR)}} - \widehat{Q}_{\mathrm{BP}}$, since these are the two best-performing tests.
[6]Note that Cases 5–9 are redundant for $k = 2$, while Case 9 is inapplicable for $k = 4$.

Table 2. Average Excess Power over Size Estimates in $\alpha \in [0.01, 0.1]$ (normally distributed errors).

k	Case	$\widehat{Q}_{\mathrm{BP}}$	$\widehat{Q}_{\mathrm{W}}$	$\widehat{Q}_{\mathrm{RS(PSR)}}$	$\widehat{Q}_{\mathrm{RS(JWW)}}$	$\mathrm{SE}(\widehat{Q}_{\mathrm{RS(PSR)}} - \widehat{Q}_{\mathrm{BP}})$
2	1A	0.177	0.1175	**0.5432**	0.02614	0.008781
	1M	0.2178	0.1467	**0.3040**	0.04448	0.009509
	2A	**0.4976**	0.3697	0.2903	0.1374	0.007898
	2M	**0.5661**	0.436	0.3545	0.1593	0.009617
	3A	**0.4913**	0.3673	0.2828	0.1362	0.009521
	3M	0.1553	0.09953	**0.4739**	0.01921	0.008049
	4A	0.3229	0.2186	**0.3569**	0.07507	0.009401
	4M	**0.3479**	0.2458	0.2533	0.07658	0.009923
4	1A	0.01352	0.01738	**0.8891**	-0.005752	0.003696
	1M	0.08318	0.05913	**0.7411**	0.01138	0.007898
	2A	**0.1960**	0.1235	0.06685	0.06474	0.007523
	2M	**0.6993**	0.5389	0.3823	0.2868	0.01013
	3A	**0.1952**	0.1212	0.07124	0.06517	0.00764
	3M	0.2616	0.1647	**0.7658**	0.08055	0.008411
	4A	0.1163	0.07299	**0.2025**	0.04139	0.008071
	4M	0.3934	0.2529	**0.4439**	0.1248	0.01029
	5A	0.1200	0.0729	**0.2084**	0.04086	0.007788
	5M	0.1172	0.07484	**0.7658**	0.02817	0.007662
	6A	0.0638	0.0425	**0.5888**	0.01883	0.007544
	6M	0.2793	0.176	**0.6528**	0.08586	0.009493
	7A	0.06394	0.03818	**0.5849**	0.0155	0.00641
	7M	0.07921	0.05655	**0.8225**	0.01043	0.006366
	8A	0.03459	0.02647	**0.6813**	0.006012	0.005859
	8M	0.1218	0.07713	**0.6131**	0.02599	0.008147
8	1A	0.002029	0.00779	**0.8993**	-0.01308	0.003
	1M	0.01417	0.01892	**0.7116**	-0.01034	0.006881
	2A	**0.05932**	0.03606	0.01531	0.005619	0.00531
	2M	**0.7604**	0.5615	0.4625	0.1739	0.009118
	3A	0.02915	0.02459	**0.4526**	-0.004934	0.007808
	3M	0.08586	0.05968	**0.7768**	0.003799	0.005964
	4A	0.04555	0.02983	**0.06457**	0.001773	0.005842
	4M	0.3546	0.2233	**0.5439**	0.05757	0.01108
	5A	0.02059	0.01748	**0.3793**	-0.002054	0.007674
	5M	0.05210	0.03914	**0.7503**	-0.00211	0.005774
	6A	0.02358	0.01723	**0.3405**	-0.001724	0.007327
	6M	0.1409	0.08909	**0.6893**	0.01353	0.007479
	7A	0.02624	0.02006	**0.3671**	0.001135	0.007275
	7M	0.04285	0.0364	**0.7474**	-0.005365	0.005948
	8A	0.02416	0.01536	**0.4117**	-0.001385	0.007494
	8M	0.1023	0.06362	**0.5857**	0.01036	0.008464
	9A	0.01975	0.01567	**0.3648**	-0.004483	0.008125
	9M	0.01707	0.01693	**0.6629**	-0.009118	0.00691

Table 3. Test Size Estimates for Nominal Size of $\alpha = 0.05$ (based on 10^4 replicates).

Error Distribution	k	BP	W	RS(PSR)	RS(JWW)
Normal		0.0464	0.0441	0.0575	0.0417
Laplace	2	0.0466	0.0467	0.3469	0.0332
Uniform		0.0484	0.0493	0.0067	0.0545
Normal		0.0439	0.0399	0.0640	0.0366
Laplace	4	0.0441	0.0439	0.4061	0.0292
Uniform		0.0470	0.0451	0.0076	0.0438
Normal		0.0415	0.0388	0.0778	0.0312
Laplace	8	0.0386	0.0406	0.4231	0.0260
Uniform		0.0430	0.0375	0.0079	0.0312

Standard error of estimates under null hypothesis that nominal size is achieved is 0.002179.

5. Discussion and Conclusion

RS(PSR) is oversized for normally distributed errors, while the other tests are undersized (especially for large k). RS(PSR) nevertheless performs better in terms of $\widehat{Q}(0.01, 0.1)$ than the other tests under numerous heteroscedastic alternatives (again, especially for large k). The simulations thus demonstrate that RS(PSR) is a powerful omnibus test for heteroscedasticity. RS(PSR) is extremely oversized when the errors are Laplace-distributed (leptokurtic) and extremely undersized when the errors are uniformly distributed (platykurtic). The size of the nonparametric test RS(JWW) is more robust but is not competitive in terms of $\widehat{Q}(0.01, 0.1)$. Further research should proceed in the following directions: (i) refine RS(PSR) to make it more robust against non-normality; (ii) identify a more powerful nonparametric test of weak stationarity to use within the RS method; (iii) develop an optimisation strategy for c, the number of 'inner' points to remove when using RS(PSR).

References

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, **47** (5), 1287–1294.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. 2nd edition. Springer, New York.

Constantine, W. and Percival, D. (2017). `fractal`*: A Fractal Time Series Modeling and Analysis Package*. R package version 2.0-4.
URL: *https://CRAN.R-project.org/package=fractal*

Goldfeld, S. M. and Quandt, R. E. (1965). Some tests for homoscedasticity. *Journal of the American Statistical Association*, **60** (310), 539–547.

Griffiths, W. E. and Surekha, K. (1986). A Monte Carlo evaluation of the power of some tests for heteroscedasticity. *Journal of Econometrics*, **31**, 219–231.

Gujarati, D. N. and Porter, D. C. (2009). *Basic Econometrics*. 5th edition. McGraw-Hill, New York.

Jin, L., Wang, S., and Wang, H. (2015). A new non-parametric stationarity test of time series in the time domain. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, **77** (5), 893–922.

Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *The Annals of Statistics*, **27** (3), 783–858.

Lloyd, C. J. (2005). Estimating test power adjusted for size. *Journal of Statistical Computation and Simulation*, **75** (11), 921–934.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, **2** (1), 49–55.

Priestley, M. B. and Subba Rao, T. (1969). A test for non-stationarity of time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, **31** (1), 140–149.

Račkauskas, A. and Zuokas, D. (2007). New tests of heteroskedasticity in linear regression model. *Lithuanian Mathematical Journal*, **47** (3), 248–265.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48** (4), 817–838.

Wooldridge, J. M. (2016). *Introductory Econometrics*. 6th edition. Cengage, Boston.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, **28** (2), 461–482.