ROBUST TESTS FOR HETEROSKEDASTICITY AND AUTOCORRELATION IN THE MULTIPLE REGRESSION MODEL

Anil K. Bera
Department of Economics
University of Illinois
Champaign, IL 61820
USA

Pin T. Ng
College of Business Administration
Northern Arizona University
Flagstaff, AZ 86011-5066
USA

Key Words and Phrases: Nonparametric statistics; Rao's score test; score function; specification tests; heteroskedasticity; autocorrelation; robustness

ABSTRACT

The standard Rao's (1948) score or Lagrange multiplier test for heteroskedasticity was originally developed assuming normality of the disturbance term [see Godfrey (1978b), and Breusch and Pagan (1979)]. Therefore, the resulting test depends heavily on the normality assumption. Koenker (1981) suggests a studentized form which is robust to nonnormality. This approach seems to be limited because of the unavailability of a general procedure that transforms a test to a robust one. Following Bickel (1978), we use a different approach to take account of nonnormality. Our tests will be based on the score function which is defined as the negative derivative of the log-density function with respect to the underlying random variable. To implement the test we use a nonparametric estimate of the score function. Our robust test for heteroskedasticity is obtained by running a regression of the product of the score function and ordinary least squares residuals on some exogenous variables which are thought to be causing the heteroskedasticity. We also use our procedure to develop a robust test for autocorrelation which can be computed by regressing the score function on the lagged ordinary least squares residuals and the independent variables. Finally, we carry out an extensive Monte Carlo study which demonstrates that our proposed tests have superior finite sample properties compared to the standard tests.

1 Introduction

Conventional model specification tests are performed with some parametric, usually the Gaussian, assumptions on the stochastic process generating a model. These parametric specification tests have the drawback of having incorrect size, suboptimal power or even being inconsistent when any of the parametric specifications of the stochastic process is incorrect. The theoretical arguments can be found in Box (1953), Tukey (1960), Bickel (1978), Koenker (1981), Im (2000), Linton and Steigerwald (2000), Machado and Santos Silva (2000), and González-Rivera and Ullah (2001) while Monte Carlo evidence is available in Bera and Jarque (1982), Davidson and MacKinnon (1983), Bera and McKenzie (1986), Linton and Steigerwald (2000), and González-Rivera and Ullah (2001). In regards to testing for autocorrelation, Evans (1992) and Ali and Sharma (1993) investigated the robustness to nonnormality of the Durbin-Watson test.

One approach to robustify a test statistic against possible departure from the Gaussian specification is to make modification to the test statistic so that it still possesses similar asymptotic distribution under a wider class of innovation densities. Koenker's (1981) modification to the standard score test for heteroskedasticity and the more recent modifications to robustify Glejser's test for heteroskedasticity by Im (2000), and Machado and Santos Silva (2000) are a few examples. This approach appears to be limited because of the unavailability of a general procedure that transforms a test to a robust one.

In this paper, we use a nonparametric estimate of the score function to develop some tests for heteroskedasticity and autocorrelation which are robust to distributional misspecifications. The importance of the score function, defined as $\psi(x) = -\log f'(x) = -\frac{f'(x)}{f(x)}$, where $f(x)$ is the probability density function of a random variable, to robust statistical procedures has been sporadically mentioned, implicitly or explicitly, throughout the past few decades [see, e.g., Hampel (1974), Bickel (1978), Koenker (1982), Joiner and Hall (1983), Manski (1984), Cox (1985), and Bera and Ng (1995)]. Numerous works have been done on nonparametric estimation of the score function, [see Stone (1975), Csörgő and Révész (1983), Manski (1984), Cox (1985), Cox and Martin (1988), and Ng (1994)]. These facilitate our development of nonparametric tests of specifications using the score function without making any explicit parametric assumption on the underlying distribution. Therefore, we expect our procedures to be immune to loss of power and incorrect size caused by distributional misspecifications.

The use of the score function in the context of model specification testing is not new. Robustifying the procedures of Anscombe (1961), and Anscombe and Tukey (1963), Bickel (1978) derives the test statistics for testing nonlinearity and heteroskedasticity which implicitly use the score function, [see also Pagan and Pak (1993)]. Linton and Steigerwald (2000) propose adaptive testing in ARCH models which also involves the score function. González-Rivera and Ullah (2001) modify the standard Rao's score (RS) tests for restrictions in a linear model and test for heteroskedasticity by using an *estimated density function* instead of an assumed parametric density. In this paper, we modify the RS test by directly estimating the *score function* nonparametrically.

Our nonparametric test for heteroskedasticity is obtained by running a regression of the product of the score function and the ordinary least squares residuals on some exogenous variables which are thought to be causing the heteroskedasticity. The nonparametric autocorrelation test is performed by regressing the score function on the lagged residuals and the independent variables, which may include lagged dependent variables. We also illustrate in the paper that when the normality assumption is true, our tests for heteroskedasticity and autocorrelation reduce to the familiar Breusch and Pagan (1979) or Godfrey (1978b) tests for heteroskedasticity and Breusch (1978) or Godfrey (1978a) tests for autocorrelation, respectively.

We perform an extensive Monte Carlo study which demonstrates that our proposed tests have superior finite sample properties compared to the standard tests when the innovation deviates from normality while still retaining comparable performances under the normal innovation.

The model and the test statistics are introduced and defined in Section 2. In Section 3, we derive the one-directional test statistics for heteroskedasticity and autocorrelation. Section 4 gives a brief review of existing score function estimation techniques and a description of the score estimator used in the Monte Carlo study. The finite sample performances of the conventional test statistics and our proposed nonparametric tests are reported in Section 5. Section 6 provides some concluding remarks.

2 The Model and the Test Statistics

2.1 The Model

In order to compare our findings with those of previous studies, we consider the following general model which incorporates various deviations from the classical linear regression model

$$\gamma(L)y_i = x_i'\beta + u_i, \qquad \delta(L)u_i = \epsilon_i, \qquad i = 1, \ldots, n \tag{1}$$

where y_i is a dependent variable, x_i is a k x 1 vector of non-stochastic explanatory variables, β is a k x 1 vector of unknown parameters, and $\gamma(L)$ and $\delta(L)$ are polynomials in the lag operator with

$$\gamma(L) = 1 - \sum_{j=1}^{m} \gamma_j L^j \qquad \text{and} \qquad \delta(L) = 1 - \sum_{j=1}^{p} \delta_j L^j.$$

The normalized innovation term is defined as $z_i = \frac{\epsilon_i}{\sigma_i}$. The innovation ϵ_i is independently distributed and has a symmetric probability density function $f_\epsilon(\epsilon_i) = \frac{1}{\sigma_i} f_z(\frac{\epsilon_i}{\sigma_i})$ with the location parameter assumed to be zero and the scale parameter taking the form $\sigma_i = \sqrt{h(v_i'\alpha)}$, in which v_i is a q x 1 vector of fixed variables having one as its first element, $\alpha' = (\alpha_1, \alpha_2')$ is a q x 1 vector of unknown parameters, and h is a known, smooth positive function with a continuous first derivative. The score function of the innovation ϵ_i is defined as

$$\psi_\epsilon(\epsilon_i) = -\frac{f_\epsilon'(\epsilon_i)}{f_\epsilon(\epsilon_i)} = -\frac{1}{\sigma_i}\frac{f_z'(z_i)}{f_z(z_i)} = \frac{1}{\sigma_i}\psi_z(\frac{\epsilon_i}{\sigma_i}). \tag{2}$$

Model (1) can be written more compactly as

$$y_i = Y_i'\gamma + x_i'\beta + u_i, \qquad u_i = U_i'\delta + \epsilon_i \tag{3}$$

where

$$\begin{array}{ccc} Y_i = (y_{i-1}, \ldots, y_{i-m})', & & U_i = (u_{i-1}, \ldots, u_{i-p})', \\ \gamma = (\gamma_1, \ldots, \gamma_m)', & \text{and} & \delta = (\delta_1, \ldots, \delta_p)' . \end{array}$$

In matrix form the model is

$$y = Y\gamma + X\beta + u = W\Gamma + u, \qquad u = U\delta + \epsilon$$

where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad Y = \begin{bmatrix} Y_1' \\ \vdots \\ Y_n' \end{bmatrix}, \quad X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

$$U = \begin{bmatrix} U_1' \\ \vdots \\ U_n' \end{bmatrix}, \quad W = \left[Y \vdots X\right] = \begin{bmatrix} W_1' \\ \vdots \\ W_n' \end{bmatrix}, \quad \Gamma = \begin{pmatrix} \gamma \\ \beta \end{pmatrix}.$$

2.2 Test Statistics

Most conventional hypotheses tests utilize the likelihood ratio (LR), Wald (W) or Rao's score (RS) principle. Each has its own appeals. The LR test is favorable when a computer package conveniently produces the constrained and unconstrained likelihoods. The Wald test is preferable when the unrestricted MLE is easier to estimate. In model specification tests, RS is the preferred principle since the null hypotheses can usually be written as restricting a subset of the parameters of interest to zero and the restricted MLE becomes the OLS estimator for the classical normal linear model.

Even though our nonparametric approach to specification tests does not lead to the OLS estimator for the restricted MLE under the null hypothesis, we will demonstrate that the RS test can still use the OLS or some other consistent estimators and specification tests can be performed conveniently by most of the popular computer packages. For this reason, we concentrate solely on deriving the RS test statistics in this paper.

Let $l_i(\theta)$ be the log-density of the ith observation, where θ is an s x 1 vector of parameters. The log-likelihood function for the n independent observations is then $l \equiv l(\theta) = \sum_{i=1}^{n} l_i(\theta)$. The hypothesis to be tested is $H_0 : h(\theta) = 0$, where $h(\theta)$ is an r x 1 vector function of θ with $r \leq s$. We denote $H(\theta) = \partial h(\theta)/\partial\theta'$ and assume that rank$(H) = r$, i.e., there are no redundant restrictions. The RS statistic is given by

$$RS = \tilde{d}'\tilde{\mathcal{I}}^{-1}\tilde{d}$$

where $d \equiv d(\theta) = \partial l/\partial\theta$ is the score vector,

$$\mathcal{I} \equiv \mathcal{I}(\theta) = Var[d(\theta)] = -E[\frac{\partial^2 l}{\partial\theta\partial\theta'}] = E[\frac{\partial l}{\partial\theta}\frac{\partial l}{\partial\theta'}]$$

is the information matrix and the '$\tilde{}$'s indicate that the quantities are evaluated at the restricted MLE of θ. Under H_0, RS is distributed as χ^2_r asymptotically.

3 Specification Tests

The usual one-directional specification tests of the model presented in (1) in Section 2.1 involve testing the following hypotheses:

1. Homoskedasticity (H): $H_0 : \alpha_2 = 0$, assuming $\delta = 0$.
2. Serial Independence (I): $H_0 : \delta = 0$, assuming $\alpha_2 = 0$.

3.1 Test for Heteroskedasticity

Breusch and Pagan (1979) derived the RS test statistic for testing the presence of heteroskedasticity under normality assumption. We provide the full derivation for the RS statistic here since the situation is somewhat different due to the nonparametric specification of the innovation distribution.

Assuming $\delta = 0$, the p.d.f. of the stochastic process specified in Section 2.1 can be written as $\frac{1}{\sigma_i} f_z(\frac{u_i}{\sigma_i})$.

We partition the vector of parameters of model (3) into

$$\theta = \begin{pmatrix} \gamma \\ \beta \\ \dots \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \dots \\ \theta_2 \end{pmatrix}.$$

The log-likelihood function is then given by

$$l(\theta) = \sum_{i=1}^{n} \left\{ \log f_z \left[\frac{u_i}{\sqrt{h(v_i'\alpha)}} \right] - \frac{1}{2}\log[h(v_i'\alpha)] \right\}.$$

The score vector under H_0 becomes

$$\begin{aligned}
\left.\frac{\partial l(\theta)}{\partial \gamma}\right|_{\tilde{\theta}} &= \sum_{i=1}^{n} - \left\{ \frac{f_z'(\frac{\tilde{u}_i}{\tilde{\sigma}})}{f_z\left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right)} \frac{1}{\tilde{\sigma}} Y_i \right\} = \sum_{i=1}^{n} \psi_z \left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right) \frac{1}{\tilde{\sigma}} Y_i = 0 \\
\left.\frac{\partial l(\theta)}{\partial \beta}\right|_{\tilde{\theta}} &= \sum_{i=1}^{n} - \left\{ \frac{f_z'(\frac{\tilde{u}_i}{\tilde{\sigma}})}{f_z\left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right)} \frac{1}{\tilde{\sigma}} x_i \right\} = \sum_{i=1}^{n} \psi_z \left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right) \frac{1}{\tilde{\sigma}} x_i = 0 \\
\left.\frac{\partial l(\theta)}{\partial \alpha}\right|_{\tilde{\theta}} &= \frac{1}{2}\sum_{i=1}^{n} \left\{ - \frac{f_z'(\frac{\tilde{u}_i}{\tilde{\sigma}})}{f_z\left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right)} \frac{\tilde{u}_i}{\tilde{\sigma}^3} h'(\tilde{\alpha}_1) v_i - \frac{h'(\tilde{\alpha}_1)}{\tilde{\sigma}^2} v_i \right\} \\
&= \frac{h'(\tilde{\alpha}_1)}{2\tilde{\sigma}^2} \sum_{i=1}^{n} v_i \left\{ \psi_z \left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right) \left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right) - 1 \right\}
\end{aligned}$$

where $\tilde{\sigma}^2 = h(\tilde{\alpha}_1)$, $\tilde{u}_i = y_i - Y_i'\tilde{\gamma} - x_i'\tilde{\beta}$, $\tilde{\alpha}_1$, $\tilde{\gamma}$ and $\tilde{\beta}$ are the restricted MLE obtained as the solutions to the above first order conditions.

If we partition the information matrix into

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix}$$

corresponding to $\theta = (\theta_1', \theta_2')'$, we can see that

$$\mathcal{I}_{12} = \mathcal{I}_{21}' = -E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2'}\right) = 0 \tag{4}$$

due to the symmetry of the p.d.f. of u_i. The lower right partition of $\mathcal{I}$ is given by

$$\mathcal{I}_{22} = Var\left[d_2(\theta)\right] = Var\left[\frac{\partial l(\theta)}{\partial \theta_2}\right].$$

Letting $c_i = \frac{1}{2}\frac{h'(v_i'\alpha)}{\sigma_i{}^2}$ and $g_i = \psi_z(\frac{u_i}{\sigma_i})(\frac{u_i}{\sigma_i}) = \psi_\epsilon(u_i)u_i$, we get

$$d_2(\theta) = \sum_{i=1}^{n} c_i v_i (g_i - 1)$$

from the first order conditions. This gives us $\mathcal{I}_{22} = \sum_{i=1}^{n} c_i^2 v_i Var(g_i) v_i'$. Denoting $\sigma_g^2 = Var(g_i)$, we have $\mathcal{I}_{22} = \sigma_g^2 \sum_{i=1}^{n} c_i^2 v_i v_i'$. We can estimate σ_g^2 by the consistent estimator

$$\hat{\sigma}_g^2 = \frac{\sum_{i=1}^{n} g_i^2}{n} - \left(\frac{\sum_{i=1}^{n} g_i}{n}\right)^2$$

and get

$$\hat{\sigma}_{\tilde{g}}^2 = \hat{\sigma}_g^2\Big|_{\tilde{\theta}} = \frac{\sum_{i=1}^n \tilde{g}_i^2}{n} - \left(\frac{\sum_{i=1}^n \tilde{g}_i}{n}\right)^2 = \frac{\sum_{i=1}^n \tilde{g}_i^2}{n} - 1$$

since $\sum_{i=1}^n \tilde{g}_i = n$ from the first order condition for α_1. Let $\tilde{g} = (\tilde{g}_1, \cdots, \tilde{g}_n)^{'}$, $V = [v_1 \cdots v_n]^{'}$ and $\mathbf{1} = (1, \cdots, 1)^{'}$. Since the information matrix is block diagonal, the RS statistics for testing $H_0 : \alpha_2 = 0$ can be written as

$$\begin{aligned} RS_H &= \tilde{d}_2' \tilde{\mathcal{I}}_{22}^{-1} \tilde{d}_2 = \frac{1}{\sigma_{\tilde{g}}^2} (\tilde{g} \ - \ 1)' V \left\{V'V\right\}^{-1} V' (\tilde{g} \ - \ 1) \\ &= \frac{1}{\sigma_{\tilde{g}}^2} \left\{ \tilde{g}'V \left(V'V\right)^{-1} V'\tilde{g} - \tilde{g}'\mathbf{1} \left(\mathbf{1}'\mathbf{1}\right)^{-1} \mathbf{1}'\tilde{g} \right\}. \end{aligned}$$

If we substitute $\hat{\sigma}_{\tilde{g}}^2$ for $\sigma_{\tilde{g}}^2$ into RS_H, we get

$$\widehat{RS}_H = \frac{1}{\hat{\sigma}_{\tilde{g}}^2} \left\{ \tilde{g}'V \left(V'V\right)^{-1} V'\tilde{g} - \tilde{g}'\mathbf{1} \left(\mathbf{1}'\mathbf{1}\right)^{-1} \mathbf{1}'\tilde{g} \right\}.$$

The RS_H test is not feasible and neither is $\widehat{RS}_H$ because the score function ψ of the innovation is unknown and, hence, prevents us from solving for the restricted MLE $\tilde{\alpha}_1$, $\tilde{\gamma}$, and $\tilde{\beta}$. To obtain a feasible version of the RS_H statistic, let $\hat{\gamma}$ and $\hat{\beta}$ be any weakly consistent estimators, e.g., the OLS estimators, for γ, and β, respectively, and $\hat{\psi}_\epsilon^*$ be a weakly consistent estimator for the true score function ψ_ϵ over the interval $[\hat{u}_{(1)}, \hat{u}_{(n)}]$. Here $\hat{u}_{(1)}$ and $\hat{u}_{(n)}$ are the extreme order statistics of the consistent residuals. Denoting $\hat{g}_i = \hat{\psi}_\epsilon^* \left(\hat{u}_i\right) \left(\hat{u}_i\right)$ and $\hat{\sigma}_{\hat{g}}^2 = \frac{\sum \hat{g}_i^2}{n} - 1$, we define our operational form of the RS statistic as

$$\widehat{\widehat{RS}}_H = \frac{1}{\hat{\sigma}_{\hat{g}}^2} \left\{ \hat{g}'V \left(V'V\right)^{-1} V'\hat{g} - \hat{g}'\mathbf{1} \left(\mathbf{1}'\mathbf{1}\right)^{-1} \mathbf{1}'\hat{g} \right\}$$

where R^2 is the centered coefficient of determination from running a regression of $\hat{g}$ on V. Following González-Rivera and Ullah (2001) and using the results in Ai (1997), it can be shown that under H_0, $\widehat{\widehat{RS}}_H \xrightarrow{D} \chi_{q-1}^2$.

Several interesting special cases can easily be derived from $\widehat{\widehat{RS}}_H$ assuming different specification for $f_\epsilon(\epsilon_i)$. For example, under the normality assumption on $f_\epsilon(\epsilon_i)$, $\psi_\epsilon(u_i) = u_i/\sigma^2$, and $\widehat{\widehat{RS}}_H - RS_{BP} \xrightarrow{p} 0$, where RS_{BP} is the RS statistic for testing heteroskedasticity in Breusch and Pagan (1979). If $f_\epsilon(\epsilon_i)$ is a double exponential distribution [Box and Tiao (1973, p.157)], $\widehat{\widehat{RS}}_H$ asymptotically becomes the Glejser's (1969) statistic which regresses $|\hat{u}_i|$on v_i, [see Pagan and Pak (1993)]. Finally, for the logistic innovation, our $\widehat{\widehat{RS}}_H$ statistic is obtained by regressing $\hat{u}_i \left(\frac{e^{\hat{u}_i}-1}{e^{\hat{u}_i}+1}\right)$on v_i. Note that the score functions for the double exponential and logistic distributions are bounded, and therefore, the latter two tests might perform better for fat tailed distributions.

3.2 Test for Serial Correlation

Given the model specified by (3) along with the assumption $\alpha = 0$, the null hypothesis for serial independence is $H_0 : \delta = 0$. Writing $\theta = \left(\theta_1, \theta_2'\right)' = \left(\sigma, \gamma', \beta', \delta'\right)'$, our model for testing serial independence can be written as

$$y_i = q_i(W_i, U_i; \theta_2) + \epsilon_i \tag{5}$$

where θ_2 is an $(m + k + p)$ x 1 vector and the ϵ_i's are I.I.D. with symmetric p.d.f. $f_\epsilon(\epsilon_i) = \frac{1}{\theta_1} f_z(\frac{\epsilon_i}{\theta_1})$, in which θ_1 is the scale parameter.

Let us denote Q as an n x $(m + k + p)$ matrix with the ith row being $\partial q_i(W_i, U_i; \theta_2)/\partial\theta_2'$, Ψ an n x 1 vector with elements $\Psi_i = \frac{1}{\theta_1}\psi_z\left(\frac{\epsilon_i}{\theta_1}\right) = \psi_\epsilon(\epsilon_i)$, and $\sigma^2_\Psi = E(\Psi_i^2)$. Proceeding as in Section 3.1, we obtain the RS_I statistic for testing $H_0 : \delta = 0$ as

$$RS_I = \frac{\tilde{\Psi}'\tilde{Q}(\tilde{Q}'\tilde{Q})^{-1}\tilde{Q}'\tilde{\Psi}}{\sigma^2_{\tilde{\Psi}}}$$

where $\sigma^2_{\tilde{\Psi}} = E(\Psi_i^2)\Big|_{\tilde{\theta}}$ and $\tilde{Q}$ is the value of Q under H_0.

Letting $\hat{\sigma}^2_{\tilde{\Psi}} = \frac{\tilde{\Psi}'\tilde{\Psi}}{n}$ be the consistent estimator for $\sigma^2_{\tilde{\Psi}}$, we have

$$\widehat{RS}_I = \frac{\tilde{\Psi}'\tilde{Q}(\tilde{Q}'\tilde{Q})^{-1}\tilde{Q}'\tilde{\Psi}}{\hat{\sigma}^2_{\tilde{\Psi}}}.$$

Similar to the test for heteroskedasticity, neither RS_I nor $\widehat{RS}_I$ is feasible. To obtain a feasible version of the RS test, let $\hat{\theta}_2$ be any weakly consistent estimator for θ_2, $\hat{\psi}^*_\epsilon$ be a weakly consistent estimator for the true score function ψ_ϵ over the interval $[\hat{\epsilon}_{(1)}, \hat{\epsilon}_{(n)}]$, $\hat{\epsilon}_i = y_i - q_i(W_i, U_i; \hat{\theta}_2)$, $\hat{\Psi}$ an n x 1 vector with elements $\hat{\Psi}_i = \hat{\psi}^*_\epsilon(\hat{\epsilon}_i)$, $\hat{Q}$ an n x $(m + k + p)$ matrix with the ith row being $\partial q_i(W_i, U_i; \hat{\theta}_2)/\partial\theta_2'$ and $\hat{\sigma}^2_{\hat{\Psi}} = \hat{\Psi}'\hat{\Psi}/n$, then the feasible RS statistic for testing serial independence in model (5) is given by

$$\widehat{\widehat{RS}}_I = \frac{\hat{\Psi}'\hat{Q}(\hat{Q}'\hat{Q})^{-1}\hat{Q}'\hat{\Psi}}{\hat{\sigma}^2_{\hat{\Psi}}} = nR^2$$

where R^2 is the uncentered coefficient of determination of regressing $\hat{\Psi}$ on $\hat{Q}$.

Notice that the n x $(m + k + p)$ matrix $\hat{Q}$ above has component $\hat{Q}_i = (Y_i', x_i', \hat{U}_i')$. This facilitates the following simpler RS statistic

$$\widehat{\widehat{RS}}_I = \frac{\hat{\Psi}'\hat{U}\left[\hat{U}'\hat{U} - \hat{U}'W\left(W'W\right)^{-1}W'\hat{U}\right]^{-1}\hat{U}'\hat{\Psi}}{\hat{\sigma}^2_{\hat{\Psi}}} = nR^2$$

where R^2 is the uncentered coefficient of determination of regressing $\hat{\Psi}$ on $\hat{U}$ and W due to the orthogonality given in the first order condition on the score vector under H_0. A well known alternative for computing the $\widehat{\widehat{RS}}_I$ statistic is to regress $\hat{\Psi}$ on $\hat{U}$ and W and test the significance of the coefficients of $\hat{U}$. Following similar arguments as in the case of heteroskedasticity, it can be shown that under serial independence, $\widehat{\widehat{RS}}_I \xrightarrow{D} \chi^2_p$.

Similar to $\widehat{\widehat{RS}}_H$, several interesting special cases can be obtained from $\widehat{\widehat{RS}}_I$. Under the normality assumption, we have $\hat{\Psi}_i = \epsilon_i$ and $\widehat{\widehat{RS}}_I - RS_{BG} \xrightarrow{p} 0$, where RS_{BG} is the RS statistic for testing autocorrelation in Breusch (1978) and Godfrey (1978a). The test can be performed by regression $\hat{\epsilon}$ on $\hat{U}$ and W. When the density of the innovation is double exponential, our test is performed by regressing $sign(\hat{\epsilon}_i)$ on $\hat{U}_i'$ and W_i'. This is similar to the sign test for randomness of a process. If the innovation has a logistic density, our $\widehat{\widehat{RS}}_I$ test is equivalent to regressing $\frac{e^{\hat{\epsilon}_i}-1}{e^{\hat{\epsilon}_i}+1}$ on $\hat{U}_i'$ and W_i'.

4 Score Function Estimation

The score function as defined in (2) plays an important role in many aspects of statistics. It can be used for data exploration purposes, for Fisher information estimation and for the construction of adaptive estimators of semiparametric econometric models in robust econometrics [see e.g. Cox and Martin (1988), Ng (1994), Bera and Ng (1995), and Steigerwald (1997)]. Here we use it to construct the nonparametric test statistics $\widehat{\widehat{RS}}_H$ and $\widehat{\widehat{RS}}_I$.

Most existing score function estimators are constructed by computing the negative logarithmic derivative of some kernel based density estimators [see e.g. Stone (1975), Manski (1984), and Cox and Martin (1988)]. Csörgő and Révész (1983) suggested a nearest-neighbor approach. Modifying the approach suggested in Cox (1985), Ng (1994) implemented an efficient algorithm to compute the smoothing spline score estimator that solved

$$\min_{\psi \in H_2[a,b]} \int (\psi^2 - 2\psi')dF_n + \lambda \int (\psi''(x))^2 dx \tag{6}$$

where $H_2[a,b] = \{\psi : \psi, \psi'$ are absolutely continuous, and $\int_a^b [\psi''(x)]^2 dx < \infty\}$. The objective function (6) is the (penalized) empirical analogue of minimizing the following mean-squared error:

$$\int (\psi - \psi_0)^2 dF_0 = \int (\psi^2 - 2\psi')dF_0 + \int \psi_0^2 dF_0 \tag{7}$$

in which ψ_0 is the unknown true score function and the equality is due to the fact that under some mild regularity conditions [see Cox (1985)]

$$\int \psi_0 \zeta dF_0 = -\int f_0'(x)\zeta(x)dx = \int \zeta' dF_0.$$

Since the second term on the right hand side of (7) is independent of ψ, minimizing the mean-squared error may focus exclusively on the first term. Minimizing (6) yields a balance between "fidelity-to-data" measured by the mean-squared error term and the smoothness represented by the second term. As in any nonparametric score function estimator, the smoothing spline score estimator has the penalty parameter λ to choose. The penalty parameter merely controls the trade-off between "fidelity-to-data" and smoothness of the estimated score function. An automatic penalty parameter choice mechanism is suggested and implemented in Ng (1994) through *robust information criteria*.

The performances of the kernel-based score estimators depend very much on using the correct kernel that reflects the underlying true distribution generating the stochastic process besides choosing the correct window width. The right choice of kernel becomes even more important for observations in the tails where density is low since few observations will appear in the tail to help smooth things out. This sensitivity to correct kernel choice is further amplified in score function estimation where higher derivatives of the density are involved [see Ng (1995)]. It is found in Ng (1995) that the smoothing spline score estimator which finds its theoretical justification from an explicit statistical decision criterion, i.e., minimizing the mean-squared error, is more robust than the *ad hoc* estimators, like the kernel-based estimators, to distribution variations. We, therefore, use it to construct our nonparametric test statistics.

Since no estimator can estimate the tails of the score function accurately, some form of trimming is needed in the tails where observations are scarce to smooth things out. Cox (1985) showed that the smoothing spline score estimator achieved uniformly weak consistency over a bounded finite support $[a_0, b_0]$ which contains the observations $x_1, \ldots, x_n$. Denoting the solution to (6) as $\hat{\psi}(x)$, the score estimator used in constructing our nonparametric statistics $\widehat{\widehat{RS}}_H$ and $\widehat{\widehat{RS}}_I$ given in Section 3 takes the form

$$\hat{\psi}^*(x) = \begin{cases} \hat{\psi}(x) & \text{if } x_{(1)} \leq x \leq x_{(n)} \\ 0 & \text{otherwise} \end{cases} . \tag{8}$$

5 Small Sample Performances

All the results on the RS statistics discussed earlier are valid only asymptotically. We would, therefore, like to study the finite sample behavior of the various statistics in this section. We are interested in the closeness of the distributions of the statistics under the null, H_0, to the asymptotic χ^2 distributions, the estimates of the probabilities of Type-I error as well as the estimated powers. The RS statistics involved in this simulation are RS_H^* [given in Godfrey

(1978b), and Breusch and Pagan (1979)], RS_I^* [given in Breusch (1978), and Godfrey (1978a)], $\widehat{\widehat{RS}}_H$ and $\widehat{\widehat{RS}}_I$.

We use the simulation models of Bera and Jarque (1982) and Bera and McKenzie (1986) so that our results can be compared with their findings. We generate the data using the model

$$y_i = \sum_{j=1}^{4} x_{ij}\beta_j + u_i$$

where $x_{i1} = 1$, x_{i2} are random variates from $N(10, 25)$, x_{i3} from the uniform $U(7.5, 12.5)$ and x_{i4} from χ^2_{10}. The regression matrix, X, remains the same from one replication to another. Serial correlated $(\overline{I})$ errors are generated by the first-order autoregressive (AR) process, $u_i = \rho u_{i-1} + \epsilon_i$, where $|\rho| < 1$. As in Bera and Jarque (1982), and Bera and McKenzie (1986), the level of autocorrelation is categorized into 'weak' and 'strong' by setting $\rho = \rho_1 = 0.3$ and $\rho = \rho_2 = 0.7$, respectively. Heteroskedasticity $(\overline{H})$ is generated by $E(\epsilon_i) = 0$ and $V(\epsilon_i) = \sigma_i^2 = 25 + \eta v_i$, where $\sqrt{v_i} \sim N(10, 25)$ and η is the parameter that determines the degree of heteroskedasticity, with $\eta = \eta_1 = 0.25$ and $\eta = \eta_2 = 0.85$ represent 'weak' and 'strong' heteroskedasticity, respectively. In order to study the robustness of the various test statistics to distributional deviations from the conventional Gaussian innovation assumption, the non-normal $(\overline{N})$ disturbances used are (1) Student's t distribution with five degrees of freedom, t_5, which represents moderately thick-tail distributions, (2) the log-normal, log, which represents asymmetric distributions, (3) the beta distribution with scale and shape parameters 7, $B(7, 7)$, which represents distributions with bounded supports, (4) the 50% normal mixture, NM, of two normal distributions, $N(-3, 1)$ and $N(3, 1)$, which represents bi-modal distributions, (5) the beta distribution with scale 3 and shape 11, $B(3, 11)$, which represents asymmetric distributions with bounded supports, and (6) the contaminated normal, CN, which is the standard normal $N(0, 1)$ with .05% contamination from $N(0, 9)$, that attempts to capture contamination in a real-life situation. All distributions are normalized to having variance 25 under H_0.

The experiments are performed for sample size $N = 25,\ 50,$ and 100. The number of replications is 250.* The Kolmogorov-Smirnov statistics for the

*A referee suggested that the simulation study be conducted with a larger number of replications. We should point out that although it is not difficult to compute our suggested tests, the whole simulation exercise is very time intensive. With 250 replications the standard error of estimating the Type-I error probabilities at the 10% nominal level will be close to $\sqrt{0.1\,(1-0.1)\,/250} \approx 0.019$, and for the power studies, the maximum standard error will be $\sqrt{0.5\,(1-0.5)\,/250} \approx 0.032$. Increasing the number of replications will not alter the conclusions drawn from our results. Nevertheless, we fully agree with the refreee's view that larger number of replications would be more desirable and would perhaps make our claims more convincing.

various RS statistics are reported in Table I.

Table I

Kolmogorov-Smirnov statistics for testing departures from the χ^2 distribution

Disturbance Distribution		Sample Size 25	50	100
$N(0,25)$	RS_H^*	.0450	.0429	.0510
	RS_I^*	.0457	.0361	.0380
	$\widehat{RS}_H$	.0734	.0504	.0288
	$\widehat{RS}_I$	.0440	.0398	.0420
t_5	RS_H^*	.0754	.1385	.1167
	RS_I^*	.0707	.0351	.0660
	$\widehat{RS}_H$	.0444	.0436	.0674
	$\widehat{RS}_I$	.0454	.0293	.0706
log	RS_H^*	.1787	$\boxed{.3005}$	$\boxed{.4767}$
	RS_I^*	.0676	.0680	.0522
	$\widehat{RS}_H$	.0440	.0394	.0371
	$\widehat{RS}_I$	.0511	.0568	.0714
$B(7,7)$	RS_H^*	.0512	.0504	.0620
	RS_I^*	.0390	.0452	.0365
	$\widehat{RS}_H$	.0399	.0653	.0472
	$\widehat{RS}_I$	.0333	.0607	.0336
NM	RS_H^*	.2372	$\boxed{.2837}$	$\boxed{.3546}$
	RS_I^*	.0453	.0242	.0470
	$\widehat{RS}_H$	.0386	.0514	.0424
	$\widehat{RS}_I$	.0333	.0509	.0276
$B(3,11)$	RS_H^*	.0393	.0817	.0379
	RS_I^*	.0721	.0539	.0457
	$\widehat{RS}_H$	.0487	.0987	.0301
	$\widehat{RS}_I$	.0947	.0685	.0496
CN	RS_H^*	.0464	.1104	$\boxed{.1685}$
	RS_I^*	.0396	.0444	.0849
	$\widehat{RS}_H$	.0447	.0416	.0387
	$\widehat{RS}_I$	.0539	.0450	.0906

The 5% critical values for the Kolmogorov-Smirnov statistic for sample sizes of 25, 50, and 100 are .2640, .1884 and .1340, respectively while the 1% critical values for 25, 50 and 100 observations are .3166, .2260, and .1608, respectively [Pearson and Hartley (1966)]. In Table I, the Kolmogorov-Smirnov statistics that are significant at the 1% level are boxed. From Table I, it is clear that no significant departure from the asymptotic χ_1^2 distribution can be concluded at either 5% or 1% levels of significance for all RS statistics under $N(0,25)$, $B(7,7)$, and $B(3,11)$. The departure from the χ^2 distribution becomes more noticeable for RS_H^* as the sample size increases when the disturbance term follows the *log*, NM or CN distributions. This is illustrated in Figure 1 for *log* and Figure 2 for the NM disturbance terms; both sample sizes equal 100.

Both figures are plots of the nonparametric adaptive kernel density estimates of RS_H^* and $\widehat{\widehat{RS}}_H$ [see Silverman (1986) for details of adaptive kernel density estimation]. We can see that RS_H^* has a thinner tail under NM and a thicker tail under *log* than the asymptotic χ_1^2 distribution. This suggests that under the null hypothesis of homoskedasticity and serial independence, the distribution of the conventional RS statistic for testing heteroskedasticity deviates away from the χ^2 distribution as the distribution of the disturbance term departs further from the normal distribution in shape while our nonparametric heteroskedasticity test statistics are more robust to these distributional deviations. From Figures 1 and 2, it is clear that at the right tail, the distributions of $\widehat{\widehat{RS}}_H$ and the χ_1^2 are very close. To maintain the correct size of a test statistic, only the right tail of its distribution matters. As we will see later in Table II, the true Type-I error probabilities of $\widehat{\widehat{RS}}_H$ are very close to the nominal level of 10%. Both the RS_I^* and $\widehat{\widehat{RS}}_I$ statistics seem to be much less sensitive to distributional deviations in the disturbance term. The estimated probabilities of Type-I error for the RS statistics are reported in Table II. The estimated probabilities are the portions of the replications for which the estimated RS statistics exceed the asymptotic 10% critical value of the χ_1^2 distribution.

From Table II, it is obvious that the Type-I error probabilities for our nonparametric test statistics, $\widehat{\widehat{RS}}_H$ and $\widehat{\widehat{RS}}_I$ are very close to the nominal 10% level under almost all sample sizes and distributions. On the other hand, the true sizes for RS_H^* could be very high. For example, when the distribution is *log*, for sample of size 100, RS_H^* rejects the true null hypothesis of homoskedasticity 54% of the times. When the distribution is t_5 or CN, RS_H^* also overly rejects, though less severely. As we have noted while discussing the implications of Figure 1, over rejection occurs since the distribution of RS_H^* has a much thicker tail when the normality assumption is violated. On the other hand, the effect of NM distribution on RS_H^* is quite the opposite. RS_H^* has a thinner tail than χ_1^2 as noted in Figure 2 resulting in very low Type-I error probabilities. The Type-I error probabilities for $\widehat{\widehat{RS}}_H$ is, in contrast, very close to the nominal significant level of 10%. As we observed in Table I that RS_I^* is not as sensitive to departures from normality as RS_H^* is and hence the deviations from the 10% Type-I error probability of RS_I^* are not as severe as those of RS_H^*. These findings are consistent with those of Bera and Jarque (1982) and Bera and McKenzie (1986).

Figure 1. Distributions of the parametric (RS_H^*) and nonparametric ($\widehat{RS}_H$) RS statistics under lognormal disturbance.

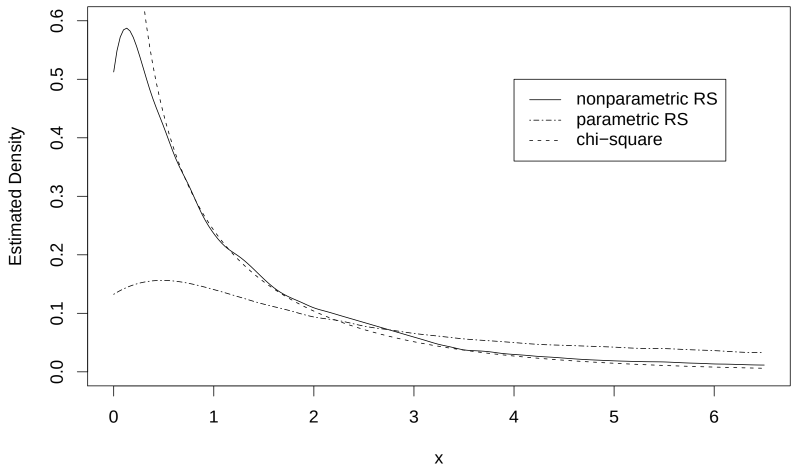

Figure 2. Distributions of the parametric (RS_H^*) and nonparametric ($\widehat{RS}_H$) RS statistics under normal-mixture disturbance.

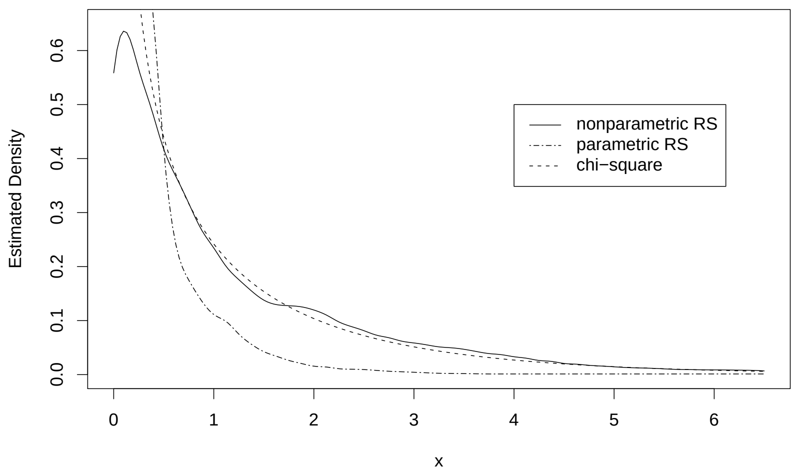

Table II
Estimated probabilities of Type-I error (nominal level 0.10)

Disturbance Distribution		Sample Size 25	50	100
$N(0,25)$	RS_H^*	.080	.116	.112
	RS_I^*	.064	.092	.096
	$\widehat{RS}_H$	.108	.128	.112
	$\widehat{RS}_I$	.076	.092	.100
t_5	RS_H^*	.108	.208	.200
	RS_I^*	.108	.088	.104
	$\widehat{RS}_H$	.108	.088	.068
	$\widehat{RS}_I$	.116	.092	.104
log	RS_H^*	.248	.388	.544
	RS_I^*	.084	.080	.060
	$\widehat{RS}_H$	.112	.068	.108
	$\widehat{RS}_I$	.100	.136	.104
$B(7,7)$	RS_H^*	.076	.072	.068
	RS_I^*	.084	.124	.100
	$\widehat{RS}_H$	.116	.100	.100
	$\widehat{RS}_I$	.072	.124	.096
NM	RS_H^*	.016	.016	.000
	RS_I^*	.144	.116	.064
	$\widehat{RS}_H$	.120	.084	.108
	$\widehat{RS}_I$	.112	.104	.084
$B(3,11)$	RS_H^*	.088	.124	.104
	RS_I^*	.076	.080	.104
	$\widehat{RS}_H$	.128	.100	.112
	$\widehat{RS}_I$	.072	.076	.100
CN	RS_H^*	.144	.204	.228
	RS_I^*	.100	.064	.140
	$\widehat{RS}_H$	.088	.092	.100
	$\widehat{RS}_I$	.092	.068	.144

Given the above results that the estimated probabilities of Type-I error for the various RS statistics are different, it is only appropriate to compare the estimate powers of the RS statistics using the simulated critical values. The $100\alpha\%$ *simulated* critical values are the $(1-\alpha)$ sample quantiles of the estimated RS statistics. The estimated powers of the RS statistics are, hence, the number of times the statistics exceed the $(1-\alpha)$ sample quantiles divided by the total number of replications. The α used in our replications is 10%. To save space, we report only the results for $N=50$ in Table III. The results for other sample sizes are qualitatively very similar and, in fact, for larger sample sizes, the better performances of our suggested nonparametric tests over their parametric counterparts are even more impressive.

Table III

Estimated power of the tests ($N = 50$, with 10% empirical significance level)

Disturbance Distributions		Alternatives: H_1							
		$\overline{H}I(\eta_1)$	$\overline{H}I(\eta_2)$	$H\overline{I}(\rho_1)$	$H\overline{I}(\rho_2)$	$\overline{HI}(\eta_1,\rho_1)$	$\overline{HI}(\eta_1,\rho_2)$	$\overline{HI}(\eta_2,\rho_1)$	$\overline{HI}(\eta_2,\rho_2)$
	RS^*_H	.592	.832	.104	.084	.548	.292	.740	.372
$N(0,25)$	RS^*_I	.112	.112	.552	.996	.576	.996	.564	.992
	$\widehat{RS}_H$	.524	.760	.100	.088	.456	.272	.688	.376
	$\widehat{RS}_I$	.116	.108	.568	.976	.568	.968	.552	.968
	RS^*_H	.448	.608	.096	.052	.388	.132	.540	.224
t_5	RS^*_I	.112	.120	.504	1.00	.524	1.00	.536	1.00
	$\widehat{RS}_H$	.432	.608	.104	.096	.396	.236	.600	.348
	$\widehat{RS}_I$	.132	.140	.504	.984	.532	.972	.544	.960
	RS^*_H	.220	.292	.084	.032	.164	.072	.268	.084
log	RS^*_I	.108	.116	.584	1.00	.572	1.00	.576	1.00
	$\widehat{RS}_H$	.600	.752	.124	.132	.472	.220	.656	.272
	$\widehat{RS}_I$	.092	.076	.748	.940	.716	.956	.656	.960
	RS^*_H	.660	.896	.100	.060	.624	.252	.828	.448
$B(7,7)$	RS^*_I	.108	.092	.528	1.00	.552	1.00	.564	1.00
	$\widehat{RS}_H$	.648	.852	.120	.088	.640	.276	.788	.424
	$\widehat{RS}_I$	.108	.092	.500	.996	.524	.992	.564	.988
	RS^*_H	.960	.996	.148	.284	.916	.500	.992	.720
NM	RS^*_I	.100	.096	.540	.984	.536	.992	.548	.992
	$\widehat{RS}_H$	.896	.956	.176	.156	.824	.352	.928	.548
	$\widehat{RS}_I$	.104	.088	.844	.980	.744	.992	.564	.988
	RS^*_H	.588	.844	.108	.092	.556	.264	.772	.404
$B(3,11)$	RS^*_I	.092	.116	.572	.992	.608	.996	.612	1.00
	$\widehat{RS}_H$	.604	.848	.108	.124	.588	.324	.784	.496
	$\widehat{RS}_I$	.116	.120	.560	.956	.572	.988	.616	.988
	RS^*_H	.396	.692	.088	.064	.400	.180	.600	.276
CN	RS^*_I	.104	.112	.524	.992	.560	.988	.548	.992
	$\widehat{RS}_H$	.488	.708	.104	.104	.448	.264	.636	.388
	$\widehat{RS}_I$	.112	.132	.524	.968	.544	.968	.564	.964

First we note that the estimated power of the parametric tests RS^*_H and RS^*_I are similar to those reported in Bera and Jarque (1982), and Bera and McKenzie (1986). Regarding the power of our nonparametric tests $\widehat{RS}_H$ and

$\widehat{\widehat{RS}}_I$, we observe that they are comparable to their parametric counterparts for $N(0, 25)$, $B(7, 7)$, $B(3, 11)$ and NM disturbances. In particular, when the disturbance distribution is normal, for which RS^*_H and RS^*_I are designed to perform best, we observe very little loss of power in using $\widehat{\widehat{RS}}_H$ and $\widehat{\widehat{RS}}_I$. On the other hand, $\widehat{\widehat{RS}}_H$ substantially outperforms its parametric counterpart when the disturbance term follows a lognormal distribution. To see the difference between the performances of RS^*_H and $\widehat{\widehat{RS}}_H$, we consider the case of lognormal distribution. RS^*_H has "optimal" power of .832 for the alternative $\overline{HI}(\eta_2)$ with normal disturbance. However, the estimated power for RS^*_H reduces to .292 when the disturbance distribution is lognormal. When we further contaminate the data by strong autocorrelation, that is under $\overline{HI}(\eta_2, \rho_2)$, the estimated power is merely .084, even less than the size of the test. The estimated powers for $\widehat{\widehat{RS}}_H$ for the above three situations are respectively .760, .752 and .272. The power reduces with gradual contamination, but not as drastically as that of RS^*_H. Note that all the distributions t_5, log, and CN, under which $\widehat{\widehat{RS}}_H$ outperforms RS^*_H, have thicker tails than the normal distribution. The $B(7, 7)$ and $B(3, 11)$ distributions, under which RS^*_H is comparable to $\widehat{\widehat{RS}}_H$, have thinner tails than the normal distribution. The NM distribution, which has the same tail behavior as the normal distribution does not deteriorate the power of RS^*_H substantially even though the distribution of RS^*_H deviates quite remarkably from the χ^2 under H_0 as we have noticed in Figure 2. The thick-tail distributions like t_5 and CN have a receding score in the tails while thin-tail distributions have a progressive score in the tails. It is exactly the thick-tail distributions that cause problems in conventional statistical methods and it is these thick-tail distributions that robust procedures are trying to deal with. The parametric RS^*_I, however, seems to be less sensitive to distributional deviation of the innovation and, hence, there are no drastic differences between RS^*_I and $\widehat{\widehat{RS}}_I$ even for severe departures from the normal distribution such as under t_5, log, and CN.

As indicated above, both the RS^*_H and $\widehat{\widehat{RS}}_H$ statistics for testing heteroskedasticity are not robust to misspecifications in serial independence. The power of both tests drops when there is severe serial correlation present in the disturbance. The effect of serial correlation is, however, more serious for RS^*_H. For instance, when the distribution is t_5, estimated power of RS^*_H reduces by .384 (= .608 − .224) as we move from $\overline{HI}(\eta_2)$ to $\overline{HI}(\eta_2, \rho_2)$. On the other hand, for $\widehat{\widehat{RS}}_H$ the power loss is .260 (= .608 − .348). This pattern is observed for almost all distributions. The power of RS^*_I and $\widehat{\widehat{RS}}_I$ are, however, more robust to violation on the maintained assumption of homoskedasticity. This is easily seen by comparing the power of RS^*_I and $\widehat{\widehat{RS}}_I$ under four sets of alternatives: (i) $H\overline{I}(\rho_1)$ and $\overline{HI}(\eta_1,\rho_1)$, (ii) $H\overline{I}(\rho_2)$ and $\overline{HI}(\eta_1,\rho_2)$, (iii) $H\overline{I}(\rho_1)$ and $\overline{HI}(\eta_2,\rho_1)$ and (iv) $H\overline{I}\,(\rho_2)$ and $\overline{HI}\,(\eta_2,\rho_2)$. Nevertheless, this suggests that some join tests or *multiple comparison procedure* in the spirit of Bera and Jarque (1982) will be able to make our tests for heteroskedasticity more robust to violation on the maintained serial independence assumption.

6 Conclusions

Our simulation results indicate that the distribution of our nonparametric *RS* statistic for testing heteroskedasticity is closer to the asymptotic χ^2 distribution under homoskedasticity and serial independence for all distributions under investigation than its parametric counterpart. The parametric *RS* statistic for testing autocorrelation is, nevertheless, much less sensitive to departure from the normality assumption and hence fares as good as its nonparametric counterpart. The estimated probabilities of Type-I error for the nonparametric *RS* statistics for testing both heteroskedasticity and autocorrelation are also much closer to the nominal 10% value. The superiority of our nonparametric *RS* test for heteroskedasticity becomes more prominent as the sample size increases and as the severity of the departure (measured roughly by the thickness in the tails) from normality increases. Therefore, we may conclude that our nonparametric test statistics are robust to distributional misspecification and will be useful in empirical work. Several extensions to our approach are possible. For instance, adopting a nonparametric conditional mean instead of the linear conditional mean model [see e.g. Lee (1992)] or even using a nonparametric conditional median specification [see e.g. Koenker, Ng and Portnoy (1994)] will further make our test statistics robust to misspecification on the conditional structural model. These extensions will be reported in future work.

ACKNOWLEDGMENTS

We are grateful to the editor and an anonymous referee for helpful suggestions that have considerably improved the paper. We would also like to thank Alok Bhargava, Roger Koenker, Paul Newbold, Steve Portnoy, Frank Vella and Bill Brown for their helpful comments. However, we retain the responsibility for any remaining errors. Part of the paper was completed while the first author was visiting CentER, Tilburg University and the second author was visiting Academia Sinica, Taiwan.

BIBLIOGRAPHY

Ai, C. (1997). "A semiparametric maximum likelihood estimator," *Econometrica*, 97, 933-963.

Ali, M.M. and Sharma, S.C. (1993). "Robustness to nonnormality of the Durbin-Watson test for autocorrelation," *Journal of Econometrics*, 57, 117-136.

Anscombe, F.J. (1961). "Examination of residuals," in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, (University of California Press, Berkeley, CA), 1-36.

Anscombe, F.J. and Tukey, J.W. (1963). "Analysis of residuals," *Technometrics*, 5, 141-160.

Bera, A. and Jarque, C. (1982). "Model specification tests: A simultaneous approach," *Journal of Econometrics*, 20, 59-82.

Bera, A. and McKenzie, C. (1986). "Alternative forms and properties of the score test," *Journal of Applied Statistics*, 13, 13-25.

Bera, A. and Ng, P. (1995). "Tests for normality using estimated score function," *Journal of Statistical Computation and Simulation*, 52, 273-287.

Bickel, P. (1978). "Using residuals robustly I: Tests for heteroscedasticity, nonlinearity," *The Annals of Statistics*, 6, 266-291.

Box, G.E.P. (1953). "Non-normality and tests on variance," *Biometrika*, 40, 318-335.

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, (Addison-Wesley, Massachusetts).

Breusch, T.S. (1978). "Testing for autocorrelation in dynamic linear models," *Australian Economic Papers*, 17, 334-55.

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

Cox, D. (1985). "A penalty method for nonparametric estimation of the logarithmic derivative of a density function," *Annals of the Institute of Statistical Mathematics*, 37, 271-288.

Cox, D. and Martin, D. (1988). "Estimation of score functions," Technical Report, University of Washington, Seattle, WA.

Csörgő, M., and Révész, P. (1983). "An N.N.-estimator for the score function," in: *Proceedings of the First Easter Conference on Model Theory*, Seminarbericht Nr.49, (der Humboldt-Universität zu Berlin, Berlin), 62-82.

Davidson, R. and MacKinnon, J. (1983). "Small sample properties of alternative forms of the Lagrange multiplier test," *Economics Letters*, 12, 269-275.

Evans, M. (1992). "Robustness of size of tests of autocorrelation and heteroscedasticity to nonnormality," *Journal of Econometrics*, 51, 7-24.

Glejser, H. (1969). "A new test for heteroscedasticity," *Journal of the American Statistical Association*, 64, 316-323.

Godfrey, L.G. (1978a). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables," *Econometrica*, 46, 1293-1301.

Godfrey, L.G. (1978b). "Testing for multiplicative heteroscedasticity," *Journal of Econometrics*, 8, 227-236.

González-Rivera, G. and Ullah, A. (2001). "Rao's score test with nonparametric density estimators," *Journal of Statistical Planning and Inference*, 97, 85-100.

Hampel, F. (1974). "The influence curve and its role in robust estimation," *Journal of The American Statistical Association*, 69, 383-393.

Im, K.S. (2000). "Robustifying Glejser test of heteroskedasticity," *Journal of Econometrics*, 97, 179-188.

Joiner, B. and Hall, D. (1983). "The ubiquitous role of $\frac{f'}{f}$ in efficient estimation of location," *The American Statistician*, 37, 128-133.

Koenker, R. (1981). "A note on studentizing a test for heteroscedasticity," *Journal of Econometrics*, 17, 107-112.

Koenker, R. (1982). "Robust methods in econometrics," *Econometric Reviews*, 1, 214-255.

Koenker, R., Ng, P. and Portnoy, S. (1994). "Quantile smoothing splines," *Biometrika*, 81, 673-680.

Lee, B. (1992). "A heteroskedasticity test robust to conditional mean misspecification," *Econometrica*, 60, 159-171.

Linton, O.B. and Steigerwald, D.G. (2000). "Adaptive testing in ARCH Models," *Econometric Reviews*, 19, 145-174.

Machado, J.S.F. and Santos Silva, J.M.C. (2000). "Glejser's test revisited," *Journal of Econometrics*, 97, 189-202.

Manski, C. (1984). "Adaptive estimation of non-linear regression models," *Econometric Reviews*, 3, 145-194.

Ng, P. (1994). "Smoothing spline score estimation," *SIAM, Journal on Scientific Computing*, 15, 1003-1025.

Ng, P. (1995). "Finite sample properties of adaptive regression estimators," *Econometric Reviews*, 14, 267-297.

Pagan, A. R. and Pak, Y. (1993). "Tests for heteroskedasticity," in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Volume 11, (North Holland, Amsterdam), 489-518.

Pearson, E.S. and Hartley, H.O. (1966). *Biometrika Tables for Statisticians*, Vol. 2, (Cambridge University Press, Cambridge, England).

Rao, C.R. (1948). "Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation," *Proceedings of the Cambridge Philosophical Society*, 44, 50-57.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, (Chapman and Hall, New York).

Steigerwald, D.G. (1997). "Uniformly adaptive estimation for models with ARMA errors," *Econometric Reviews*, 16, 393-409.

Stone, C. (1975). "Adaptive maximum likelihood estimators of a location parameter," *The Annals of Statistics*, 3, 267-284.

Tukey, J.W. (1960). "A survey of sampling from contaminated distributions," in: I. Olkin, ed., *Contributions to Probability and Statistics*, (Stanford University Press, Stanford, CA).