ROBUST TESTS FOR HETEROSKEDASTICITY AND AUTOCORRELATION USING SCORE FUNCTION

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

Pin T. Ng
Department of Economics
University of Houston
Houston, TX 77204-5882
USA

Key Words and Phrases: Nonparametric statisitics; score function; specification tests; heteroskedasticity; autocorrelation; robustness

ABSTRACT

The standard Lagrange multiplier test for heteroskedasticity was originally developed assuming normality of the disturbance term [see Godfrey (1978b), and Breush 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 sizes, suboptimal power or even being inconsistent when any of the parametric specifications of the stochastic process is incorrect, [see Box (1953), Tukey (1960), Bickel (1978) and Koenker (1981) for theoretical arguments, and Bera and Jarque (1982), Bera and McKenzie (1986), and Davidson and MacKinnon (1983) for Monte Carlo evidence]. In this paper, we use a nonparametric estimate of 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 (1973, 1974), Bickel (1978), Koenker (1982), Joiner and Hall (1983), Manski (1984), and Cox (1985)]. Only during the past decade, numerous works were 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, 1991)]. 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 powers and incorrect sizes 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 (1982) derives the test statistics for testing nonlinearity and heteroskedasticity which implicitly use the score function, [see also Pagan and Pak (1991)]. In this paper, we follow the Lagrange multiplier test procedure and derive the test statistics which turn out to be functions of the score function.

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 show in the paper that when 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 retain 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.

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

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

$$\delta(L)\, u_i \;=\; \epsilon_i \tag{2}$$

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

$$\gamma(L) \;=\; 1 \;-\; \sum_{j=1}^{m} \gamma_j L^j$$

$$\delta(L) \;=\; 1 \;-\; \sum_{j=1}^{p} \delta_j L^j \;.$$

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 continuous first derivative. The score function of the innovation ϵ_i is defined as

$$\psi_\epsilon(\epsilon_i) \;=\; -\frac{f_\epsilon'(\epsilon_i)}{f_\epsilon(\epsilon_i)} \;=\; -\frac{1}{\sigma_i}\frac{f_z'(z_i)}{f_z(z_i)} \;=\; \frac{1}{\sigma_i}\psi_z(\frac{\epsilon_i}{\sigma_i}). \tag{3}$$

Model (1) and (2) can be written more compactly as

$$\begin{aligned} y_i &= Y_i'\gamma + x_i'\beta + u_i && (4) \\ u_i &= U_i'\delta + \epsilon_i && (5) \end{aligned}$$

where

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

In matrix form the model is

$$\begin{aligned} y &= Y\gamma + X\beta + u = W\Gamma + u \\ u &= U\delta + \epsilon \end{aligned}$$

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 testings utilize the likelihood ratio (LR), Wald (W), or Lagrange multiplier (LM) 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, LM 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 LM test can still use the OLS or some other consistent estimators and specification tests can be performed conveniently through most of the popular computer packages. For this reason, we concentrate solely on deriving the LM test statistics in this paper.

Let $l_i(\theta)$ be the log-density of the ith observation, where θ is a sx1 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:

$$H_0 : \quad h(\theta) = 0 .$$

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 LM statistic is given by

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

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 '~'s indicate that the quantities are evaluated at the restricted MLE of θ. Under H_0, LM is distributed as χ^2_r asymptotically.

3 Specification Tests

The usual one-directional specification tests of the model given by (1) and (2) 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 LM test statistic for testing the presence of heteroskedasticity under normality assumption. Here we provide the full derivation for the LM statistic 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 shall partition the vector of parameters of model (4) and (5) 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

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

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

$$\begin{aligned} \mathcal{I}_{12} &= \mathcal{I}_{21}' = -E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2'} \right) \\ &= E\left\{ \frac{1}{2} \sum_{i=1}^{n} \frac{h'(v_i'\alpha)}{\sigma_i^2} \left[\psi_z(\frac{u_i}{\sigma_i})(\frac{1}{\sigma_i}) W_i' + (\frac{u_i}{\sigma_i^2}) \psi_z'(\frac{u_i}{\sigma_i}) W_i' \right] v_i' \right\} . \quad (6) \end{aligned}$$

With the assumption of a symmetric p.d.f. for u_i, $\mathcal{I}_{12} = \mathcal{I}_{21}' = 0$ due to the fact that both terms in (6) are odd functions. The lower right partition of $\mathcal{I}$ is given by

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

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

$$\tilde{g} = \begin{pmatrix} \tilde{g}_1 \\ \vdots \\ \tilde{g}_n \end{pmatrix}, \quad V = \begin{bmatrix} v_1' \\ \vdots \\ v_n' \end{bmatrix}, \quad \mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Since the information matrix is block diagonal, the LM statistics for testing

$$H_0: \quad \left[0 \,\vdots\, I\right] \begin{bmatrix} \alpha_1 \\ \cdots \\ \alpha_2 \end{bmatrix} = 0$$

can be written as

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

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

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

The LM_H test is not feasible and neither is $\widehat{LM}_H$ 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 LM_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 $\hat{g}_i = \hat{\psi}^*_\epsilon(\hat{u}_i)(\hat{u}_i)$ and $\hat{\sigma}^2_{\hat{g}} = \frac{\sum \hat{g}_i^2}{n} - 1$, we define our operational form of the LM statistic as

$$\begin{aligned}\widehat{\widehat{LM}}_H &= \frac{1}{\hat{\sigma}^2_{\hat{g}}}\left\{\hat{g}'V\left(V'V\right)^{-1}V'\hat{g} \;-\; \hat{g}'\mathbf{1}\left(\mathbf{1}'\mathbf{1}\right)^{-1}\mathbf{1}'\hat{g}\right\} \\ &= nR^2\end{aligned}$$

where R^2 is the centered coefficient of determination from running a regression of $\hat{g}$ on V.

We now demonstrate that under the null hypothesis, $\widehat{\widehat{LM}}_H$ is asymptotically distributed as χ^2_{q-1}. Since LM_H is the standard Lagrange multiplier statistic, under H_0, $LM_H \xrightarrow{D} \chi^2_{q-1}$. Under homoskedasticity, we are in an IID set up and hence $\hat{\sigma}^2_{\tilde{g}} - \sigma^2_{\tilde{g}} = o_p(1)$. Hence, under the null, LM_H and $\widehat{LM}_H$ will have the same asymptotic distribution. Next we show the asymptotic equivalence of $\widehat{LM}_H$ and $\widehat{\widehat{LM}}_H$. First, we note that, under H_0, $\hat{u}_i - \tilde{u}_i = o_p(1)$. Since ψ_ϵ is a continuous function, $\psi_\epsilon(\hat{u}_i) - \psi_\epsilon(\tilde{u}_i) = o_p(1)$. With $\hat{\psi}^*_\epsilon$ being consistent over $[\hat{u}_{(1)}, \hat{u}_{(n)}]$, $\hat{g}_i - \tilde{g}_i = o_p(1)$, and therefore $\hat{\sigma}^2_{\hat{g}} - \hat{\sigma}^2_{\tilde{g}} = o_p(1)$. Next we consider the numerators of $\widehat{LM}_H$ and $\widehat{\widehat{LM}}_H$. These numerators are based on the OLS regression of, respectively $\tilde{g}$ and $\hat{g}$ on V. Let us denote $\tilde{\eta} = (V'V)^{-1}V'\tilde{g}$ and $\hat{\eta} = (V'V)^{-1}V'\hat{g}$. Now denoting $d = \hat{g} - \tilde{g}$, we have

$$\begin{aligned}\hat{\eta} - \tilde{\eta} &= \left(\frac{V'V}{n}\right)^{-1}\left(\frac{V'd}{n}\right) \\ &= \left(\frac{V'V}{n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} v_i d_i\end{aligned}$$

Cox (1985,p.276) showed that $|d_i| = |\hat{g}_i - \tilde{g}_i| = O_p(n^{-\delta})$ for $0 < \delta < \frac{1}{2}$. Therefore, $\frac{1}{n^{1-\delta}}\sum_{i=1}^{n}|d_i| = O_p(1)$. Suppose $|v_1|, \ldots, |v_n|$ are bounded by $m < \infty$, then we can write

$$\frac{1}{n}\left|\sum_{i=1}^{n} v_i d_i\right| \;\leq\; \frac{1}{n}\sum_{i=1}^{n}|v_i d_i|$$

$$
\begin{aligned}
&\leq \frac{m}{n}\sum_{i=1}^{n}|d_i| \\
&= \frac{m}{n^{\delta}}\frac{1}{n^{1-\delta}}\sum_{i=1}^{n}|d_i| \\
&= o_p(1)
\end{aligned}
$$

This establishes that $\widehat{LM}_H$ and $\widehat{\widehat{LM}}_H$ are asymptotically equivalent, and hence, under H_0, $\widehat{\widehat{LM}}_H \xrightarrow{D} \chi^2_{q-1}$.

Several interesting special cases can easily be derived from $\widehat{\widehat{LM}}_H$ 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 $\widehat{\widehat{LM}}_H - LM_{BP} \xrightarrow{p} 0$, where LM_{BP} is the LM 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)], $\widehat{\widehat{LM}}_H$ asymptotically becomes the Glesjer's (1969) statistic which regresses $|\hat{u}_i|$ on v_i, [see Pagan and Pak (1991)]. Finally, for the logistic innovation, our $\widehat{\widehat{LM}}_H$ 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 (4) and (5) along with the assumption $\alpha = 0$, the null hypothesis for no serial independence is

$$H_0: \quad \delta = 0.$$

Writing

$$\theta = \begin{pmatrix} \theta_1 \\ \cdots \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \sigma \\ \cdots \\ \gamma \\ \beta \\ \delta \end{pmatrix},$$

our model for testing serial independence can be written as

$$y_i = q_i(W_i, U_i; \theta_2) + \epsilon_i \tag{7}$$

where θ_2 is a $(m+k+p)$ 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.

The log-likelihood function is

$$l(\theta) = \sum_{i=1}^{n}\left[\log f_z\left(\frac{\epsilon_i}{\theta_1}\right) - \log\theta_1\right]$$

and the first order conditions for the restricted MLE are

$$\begin{aligned}
\left.\frac{\partial l}{\partial \gamma}\right|_{\tilde{\theta}} &= \sum_{i=1}^{n} \psi_z\left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right)\frac{1}{\tilde{\sigma}}Y_i = 0 \\
\left.\frac{\partial l}{\partial \beta}\right|_{\tilde{\theta}} &= \sum_{i=1}^{n} \psi_z\left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right)\frac{1}{\tilde{\sigma}}x_i = 0 \\
\left.\frac{\partial l}{\partial \delta}\right|_{\tilde{\theta}} &= \sum_{i=1}^{n} \psi_z\left(\frac{\tilde{u}_i}{\tilde{\sigma}}\right)\frac{1}{\tilde{\sigma}}\tilde{U}_i
\end{aligned}$$

where the '~'s again denote quantities evaluated at the restricted MLE, $\tilde{u}_i = y_i - Y_i'\tilde{\gamma} - x_i'\tilde{\beta}$, and $\tilde{U}_i = (\tilde{u}_{i-1}, \ldots, \tilde{u}_{i-p})'$.

With the symmetry assumption on $f_\epsilon(\epsilon_i)$, as before, it can be easily proved that

$$E(\partial^2 l(\theta)/\partial\theta_2\partial\theta_1) = 0 .$$

We, therefore, only need to evaluate d_2 and $\mathcal{I}_{22}$ if we are testing for restrictions on θ_2. Denoting Q as a n x $(m+k+p)$ matrix with the ith row being $\partial q_i(W_i, U_i; \theta_2)/\partial\theta_2'$ and Ψ a n x 1 vector with elements $\Psi_i = \frac{1}{\theta_1}\psi_z\left(\frac{\epsilon_i}{\theta_1}\right) = \psi_\epsilon(\epsilon_i)$, we have

$$\begin{aligned}
d_2 &= -\sum_{i=1}^{n} \frac{f_z'\left(\frac{\epsilon_i}{\theta_1}\right)}{f_z\left(\frac{\epsilon_i}{\theta_1}\right)}\frac{1}{\theta_1}Q_i \\
&= Q'\Psi
\end{aligned}$$

and

$$\begin{aligned}
E\left[d_2 d_2'\right] &= Q'\left(E\Psi\Psi'\right)Q \\
&= \sigma_\Psi^2 Q'Q
\end{aligned}$$

where $\sigma_\Psi^2 = E(\Psi_i)^2$. The LM_I statistic for testing $H_0: \delta = 0$ is given by

$$LM_I = \frac{\tilde{\Psi}'\tilde{Q}(\tilde{Q}'\tilde{Q})^{-1}\tilde{Q}'\tilde{\Psi}}{\sigma_{\tilde{\Psi}}^2}$$

where $\sigma_{\tilde{\Psi}}^2 = E(\Psi_i^2)\big|_{\tilde{\theta}}$.

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

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

Similar to the test for heteroskedasticity, neither LM_I nor $\widehat{LM}_I$ is feasible. To obtain a feasible version of the LM 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 $[\hat{\epsilon}_{(1)}, \hat{\epsilon}_{(n)}]$, $\hat{\epsilon}_i = y_i - q_i(W_i, U_i; \hat{\theta}_2)$, $\hat{\Psi}$ a n x 1 vector with elements $\hat{\Psi}_i = \hat{\psi}^*_\epsilon(\hat{\epsilon}_i)$, $\hat{Q}$ a 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 LM statistic for testing serial independence in model (7) is given by

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

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 LM statistic.

$$\widehat{\widehat{LM}}_I = \frac{\hat{\Psi}'\hat{U}\left[\hat{U}'\hat{U} - \hat{U}'W(W'W)^{-1}W'\hat{U}\right]^{-1}\hat{U}'\hat{\Psi}}{\hat{\sigma}^2_{\hat{\Psi}}} = nR^2 \tag{8}$$

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 $\widehat{\widehat{LM}}_I$ 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, we can show that under serial independence, $\widehat{\widehat{LM}}_I \xrightarrow{D} \chi^2_p$.

As in the case of $\widehat{\widehat{LM}}_H$, several interesting special cases can be obtained from $\widehat{\widehat{LM}}_I$. Under normality assumption, we have $\hat{\Psi}_i = \epsilon_i$ and $\widehat{\widehat{LM}}_I - LM_{BG} \xrightarrow{p} 0$, where LM_{BG} is the LM 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 $\widehat{\widehat{LM}}_I$ 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 (3) plays an important role in many aspects of statistics. It can be used for data exploration purpose, 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) and Ng (1994)]. Here we use it to construct the nonparametric test statistics $\widehat{\widehat{LM}}_H$ and $\widehat{\widehat{LM}}_I$.

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

$$\min_{\psi \in H_2[a,b]} \int (\psi^2 - 2\psi')dF_n + \lambda \int (\psi''(x))^2 dx \tag{9}$$

where $H_2[a,b] = \{\psi : \psi, \psi'$ are absolutely continuous, and $\int_a^b [\psi''(x)]^2 dx < \infty\}$. The objective function (9) is the (penalized) empirical analogue of minimizing the following mean-squared error:

$$\int (\psi - \psi_0)^2 dF_0 = \int (\psi^2 - 2\psi')dF_0 + \int \psi_0^2 dF_0 \tag{10}$$

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 (10) is independent of ψ, minimizing the mean-squared error may focus exclusively on the first term. Minimizing (9) 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 tradeoff 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* [see Ng (1991) for a FORTRAN source codes].

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 (9) as $\hat{\psi}(x)$, the score estimator used in constructing our nonparametric statistics $\widehat{\widehat{LM}}_H$ and $\widehat{\widehat{LM}}_I$ given in Section 3 takes the form

$$\hat{\psi}^*(x) \;=\; \begin{cases} \hat{\psi}(x) & \text{if } x_{(1)} \;\le\; x \;\le\; x_{(n)} \\ 0 & \text{otherwise} \end{cases} . \tag{11}$$

5 Small Sample Performances

All the results on the LM 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 LM statistics involved in this simulation are LM_H^* [given in Godfrey (1978b), and Breusch and Pagan (1979)], LM_I^* [given in Breusch (1978), and Godfrey (1978a)], $\widehat{\widehat{LM}}_H$ and $\widehat{\widehat{LM}}_I$. For the LM statistics, the closeness of the distributions under the null to the asymptotic χ^2 distributions are measured by the Kolmogorov-Smirnov statistics, the estimated probabilities of Type-I errors are measured by the portion of rejections in the replications when the asymptotic χ^2 significant values are used, and the estimated powers are measured by the number of times the test statistics exceeded the corresponding empirical significant points divided by the total number of replications.

We are using the simulation models of Bera and Jarque (1982) and Bera and McKenzie (1986) so that our results can be compared with their prior findings. The linear regression model is given by

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

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, remain 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 $(\overline{H})$ are 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) the Student's t distribution with five degrees of freedom, t_5, which represent moderately thick-tail distributions, (2) the log-normal, log, which represent asymmetric distributions, (3) the beta distribution with scale and shape parameters 7, $B(7,7)$, which represent 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. Figure 1 presents the score functions of all the above distributions. Notice from Figure 1 that distributions with thicker tails than the normal have receding score in the tails while those with thinner tails than the normal have progressive score in the tails.

The experiments are performed for sample size $N = 25$, 50, and 100. The number of replication is 250. The Komogorov-Smirnov statistics for the various LM statistics are reported in TABLE I.

TABLE I

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

The 5% critical values for the Kolmogorov-Smirnov statistic for the sample sizes of 25, 50, and100 are .2640, .1884 and .1340 respectively while the 1% critical values for 25, 50 and100 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% level of significance for all LM statistics under $N(0,25)$, $B(7,7)$, and $B(3,11)$. The departure from the χ^2 distribution becomes more noticeable for LM_H^* as the sample size gets bigger when the disturbance term follows the *log*, NM or CN distributions. This is illustrated in Figure 2 for *log* and Figure 3 for the NM disturbance terms; both sample sizes equal 100. Both figures are plots of the nonparametric adaptive kernel density estimates of LM_H^* and $\widehat{\widehat{LM}}_H$ [see Silverman (1986) for details of adaptive kernel density estimation]. We can see that LM_H^* has thinner tail under NM and 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 LM 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 2 and 3, it is clear that at the tails, the distributions of $\widehat{\widehat{LM}}_H$ and the χ_1^2 are very close. To maintain the correct size of a test statistic, only the tail of its distribution matters. As we will see later in TABLE II, the true Type-I error probabilities of $\widehat{\widehat{LM}}_H$ are very close to the nominal level of 10%. Both the LM_I^* and $\widehat{\widehat{LM}}_I$ statistics seem to be much less sensitive to distributional deviations in the disturbance term.

The estimated probabilities of Type-I error for the LM statistics are reported in TABLE II. The estimated probabilities are the portions of the replications for which the estimated LM statistics exceed the asymptotic 10% critical values of the χ_1^2 distributions. Since the number of replication is 250, the standard errors of the estimated probabilities of Type-I error is no bigger than $\sqrt{0.5(1-0.5)/250} \simeq 0.032$.

TABLE II

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

From TABLE II, it is obvious that the Type-I error probabilities for our non-parametric test statistics, $\widehat{\widehat{LM}}_H$ and $\widehat{\widehat{LM}}_I$ are very close to the nominal 10% level under almost all sample sizes and distributions. On the other hand, the true sizes for LM_H^* could be very high. For example, when the distribution is *log*, for sample of size 100, LM_H^* rejects the true null hypothesis of homoskedasticity 54% of the times. When the distribution is t_5 or CN, LM_H^* also overly rejects, though less severely. As we have noted while discussing the implications of Figure 2, over rejection occurs since the distribution of LM_H^* has much thicker tail when the normality assumption is violated. On the other hand, the effect of NM distribution on LM_H^* is quite the opposite. LM_H^* has thinner tail than χ_1^2 as noted in Figure 3 resulting in very low Type-I error probabilities. The Type-I error probabilities for $\widehat{LM}_H$ is, in contrast, very close to the nominal significant level of 10%.

As we observed in TABLE I that LM_I^* is not as sensitive to departures from normality as LM_H^* is and hence the deviations from the 10% Type-I error probability of LM_I^* are not as severe as those of LM_H^*. These findings are consistent with those of Bera and Jarque (1982) and Bera and McKenzie (1986), in which the LM_H^* and LM_I^* tests have incorrect Type-I error probabilities under *log* and t_5 when the asymptotic critical values of the χ^2 distribution are used.

Given the above results that the estimated probabilities of Type-I error for the various LM statistics are different, it is only appropriate to compare the estimate powers of the LM statistics using the simulated critical values. The $100\alpha\%$ simulated critical values are the $(1 - \alpha)$ sample quantiles of the estimated LM statistics. The estimated powers of the LM 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%. The standard errors of the estimated powers are again ≤ 0.032. The estimated powers for $N = 50$ and 100 are presented in TABLE III and IV respectively.

TABLE III

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

TABLE IV

Disturbance Distributions		Alternatives: H_1 $\overline{HI}(\eta_1)$	$\overline{HI}(\eta_2)$	$H\overline{I}(\rho_1)$	$H\overline{I}(\rho_2)$	$\overline{HI}(\eta_1,\rho_1)$	$\overline{HI}(\eta_1,\rho_2)$	$\overline{HI}(\eta_2,\rho_1)$	$\overline{HI}(\eta_2,\rho_2)$
$N(0,25)$	LM_H^*	.840	.988	.092	.060	.804	.412	.968	.664
	LM_I^*	.124	.132	.864	1.00	.848	1.00	.848	1.00
	$\widehat{LM}_H$	.808	.976	.100	.072	.784	.408	.952	.640
	$\widehat{LM}_I$	.120	.132	.852	1.00	.848	1.00	.848	1.00
t_5	LM_H^*	.688	.916	.080	.040	.636	.300	.876	.492
	LM_I^*	.080	.068	.828	1.00	.860	1.00	.876	1.00
	$\widehat{LM}_H$	.764	.952	.144	.116	.700	.448	.884	.648
	$\widehat{LM}_I$	.084	.084	.828	1.00	.876	.992	.896	.996
log	LM_H^*	.256	.364	.088	.008	.184	.052	.300	.080
	LM_I^*	.108	.112	.912	1.00	.904	1.00	.900	1.00
	$\widehat{LM}_H$	.880	.972	.136	.124	.764	.368	.928	.540
	$\widehat{LM}_I$	.116	.080	.988	1.00	.996	.992	.996	1.00
$B(7,7)$	LM_H^*	.928	.996	.104	.100	.896	.532	.980	.816
	LM_I^*	.120	.120	.852	1.00	.832	1.00	.844	1.00
	$\widehat{LM}_H$	.900	.992	.096	.088	.848	.492	.964	.748
	$\widehat{LM}_I$	.116	.120	.848	.996	.828	1.00	.848	1.00
NM	LM_H^*	1.00	1.00	.212	.308	1.00	.792	1.00	.944
	LM_I^*	.104	.112	.908	1.00	.892	1.00	.896	1.00
	$\widehat{LM}_H$	.992	.996	.084	.096	.960	.512	.988	.780
	$\widehat{LM}_I$	.112	.092	1.00	1.00	.984	1.00	.972	1.00
$B(3,11)$	LM_H^*	.856	.984	.080	.052	.792	.444	.968	.712
	LM_I^*	.112	.108	.884	1.00	.864	1.00	.876	1.00
	$\widehat{LM}_H$	.884	.984	.064	.080	.792	.448	.960	.712
	$\widehat{LM}_I$	.116	.092	.900	1.00	.880	.996	.880	1.00
CN	LM_H^*	.596	.828	.108	.032	.548	.220	.780	.376
	LM_I^*	.068	.072	.760	1.00	.752	1.00	.752	1.00
	$\widehat{LM}_H$	.648	.896	.120	.108	.588	.340	.848	.496
	$\widehat{LM}_I$	.088	.092	.776	1.00	.784	.996	.800	1.00

First we note that the estimated powers of the parametric tests LM_H^* and LM_I^* are similar to those reported in Bera and Jarque (1982), and Bera and McKenzie (1986). Regarding the powers of our nonparametric tests $\widehat{LM}_H$ and $\widehat{LM}_I$, 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 LM_H^* and LM_I^* are designed to perform best, we observe very little loss of power in using $\widehat{\widehat{LM}}_H$ and $\widehat{\widehat{LM}}_I$. On the other hand, $\widehat{\widehat{LM}}_H$ substantially outperform its parametric counterpart when the disturbance term follows a lognormal distribution. To see the difference between the performances of LM_H^* and $\widehat{\widehat{LM}}_H$, we consider the case of lognormal distribution with sample size 50. LM_H^* has "optimal" power of .832 for the alternative $\overline{HI}(\eta_2)$ with normal disturbance. However, the estimated power for LM_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 $\widehat{\widehat{LM}}_H$ for the above three situations are respectively .760, .752 and .272. The power do reduces with gradual contamination, but not as drastically as that of LM_H^*. For the t_5 and CN disturbances, the advantage of the nonparametric $\widehat{\widehat{LM}}_H$ becomes more eminant as the sample size gets bigger, under which the nonparametric efficiency begins to show up. Note that all the distributions t_5, *log*, and CN, under which $\widehat{\widehat{LM}}_H$ outperforms LM_H^*, have thicker tails than the normal distribution. The $B(7,7)$ and $B(3,11)$ distributions, under which LM_H^* is comparable to $\widehat{\widehat{LM}}_H$, 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 LM_H^* substantially even though the distribution of LM_H^* deviates quite remarkably from the χ^2 under H_0 as we noticed in Figure 3. As we noted in Figure 1, the thick-tails distributions like t_5 and CN have receding score in the tails while thin-tails distributions have progressive score in the tails. It is exactly the thick-tails distributions that cause problems in conventional statistical methods and it is these thick-tails distributions that robust procedures are trying to deal with.

The parametric LM_I^*, however, seems to be less sensitive to distributional deviation of the innovation and, hence, there are no drastic differences between LM_I^* and $\widehat{\widehat{LM}}_I$ even for severe departures from the normal distribution such as under t_5 *log*, and CN.

As was indicated above, both the LM_H^* and $\widehat{\widehat{LM}}_H$ statistics for testing heteroskedasticity are not robust to misspecifications in serial independence. The power of both tests drop when there are severe serial correlations present in the disturbances. The effect of serial correlation is, however, more serious for LM_H^*. For instance, when the sample size is 100 and the distribution is t_5, estimated power of LM_H^* reduces by .424 (= .916 − .492) as we move from $\overline{HI}(\eta_2)$ to $\overline{HI}(\eta_2, \rho_2)$. On the other hand, for $\widehat{\widehat{LM}}_H$ the power loss is .304 (= .952 − .648). This pattern is observed for almost all distributions. The powers of LM_I^* and $\widehat{\widehat{LM}}_I$ are, however, more robust to violation on the maintained assumption of homoskedasticity. This is easily seen by looking at the powers of LM_I^* and $\widehat{\widehat{LM}}_I$ under three sets of alternatives: (i) $H\overline{I}(\rho_1)$ and $H\overline{I}(\rho_2)$; (ii) $\overline{HI}(\eta_1, \rho_1)$ and $\overline{HI}(\eta_1, \rho_2)$, and (iii) $\overline{HI}(\eta_2, \rho_1)$ and $\overline{HI}(\eta_2, \rho_2)$. Nevertheless, this suggests that some join tests or *Multiple Comparison Procedure* in the same 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. Furthermore by 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.

Our simulation results indicate that the distribution of our nonparametric LM statistic for testing heteroskedasticity are closer to the asymptotic χ^2 distribution under homoskedasticity and serial independence for all distributions under investigation than its parametric counterpart. The parametric LM 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 LM statistics for testing both heteroskedasticity and autocorrelation are also much closer to the nominal 10% value. The superiority of our nonparametric LM 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.

ACKNOWLEDGEMENTS

The authors would like to thank Alok Bhargava, Roger Koenker, Paul Newbold, Steve Portnoy, Frank Vella and, especially, Bill Brown for their helpful comments. Part of the paper was completed while the first author was visiting CentER, Tilburg University. The second author would also like to acknowledge financial support from the President's Research and Scholarship Fund of University of Houston.

BIBLIOGRAPHY

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.

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, Mass.

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

Breusch, T.S. and Pagen, 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.

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

Glesjer, 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.

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 Review*, 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.

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

Ng, P. (1991). "Computing smoothing spline score estimator," Discussion paper (University of Houston, Houston, TX).

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*, forthcoming.

Pagan, A. R. and Pak, Y. (1991). "Tests for heteroskedasticity," Working paper (University of Rochester, New York, NY).

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

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, New York.

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).