Outlier detection tests based on martingale estimating equations for stochastic processes

R.M. Huggins

Department of Statistics, La Trobe University, Bundoora, Vic. 3083, Australia

Abstract

An outlier detection test related to a robustified score test is proposed and compared with the sign test and other tests based on functions of estimated residuals. Examples of an autoregressive process and a regression model with autoregressive errors are presented to illustrate the techniques.

Keywords: Outlier detection; Estimating function

1. Introduction

The use of residuals for diagnostic and outlier detection tests is well known in regression analysis and the use of sums of the signs of residuals in outlier detection tests has been studied by Brown (1975), Brown and Kildea (1979), see also David (1962). More recently this approach has been extended to stochastic processes by Huggins (1989). These tests are based on noting that in the presence of asymmetric contamination the sum of the signs of residuals resulting from a non-robust estimating procedure becomes large which enables the construction of outlier detection tests. However, as revealed in simulations below, the power of tests based on the signs of residuals can be poor and in particular in small samples the discrete nature of the sign test can cause problems in determining the appropriate size of the test. Further, the sign test is no longer strictly non-parametric as its asymptotic variance depends on assumptions about the error distribution. This motivates us to search for tests of increased power by considering more informative functions of the residuals.

More generally, we apply our results to test if an easily calculated estimator, such as the least squares estimator, is the solution of a set of robust estimating equations which guard against outliers. Typically these robust estimating equations involve vectors of weighted sums of functions of standardised residuals, see for example Denby and Martin (1979), Martin (1979, 1980), Bustos (1982), Basawa et al. (1985), Godambe (1985), Martin and Yohai (1985, 1986), and Kulkarni and Heyde (1987), amongst others. Out test statistic in this setting is akin to the robustified score statistic

of Basawa et al. (1985) which is preferred to the direct comparison of the least squares and a robust estimate, a type of Wald statistic to which our tests is asymptotically equivalent, for computational reasons. This approach is in line with the estimating equation approach of Godambe and Heyde (1987) which focuses on the estimating equations rather than the resulting estimator. A similar philosophy is evident in Basawa (1985) and Basawa et al. (1985) where a preliminary estimator is adjusted by the estimating equations to provide a robust estimator. Unlike the leave-k-out diagnostics of Bruce and Martin (1989) and the tests of Fox (1972) our concern is less with the identification of particular outliers and more with establishing the presence of outliers that influence the parameter estimates.

The properties of the tests result from the derivation of the asymptotic distribution of functions of the estimated residuals from fitting a model to a stochastic process, where the residuals are standardised using a robust estimate of scale. A robust estimate of scale is used to prevent the masking of outliers by inflated estimates of scale. The procedure is described in Section 2, some examples are given in Section 3 and some simulations are discussed in Section 4.

2. The test statistics

Let $\{X_n, \mathscr{F}_n; n \geq 1\}$ be a stochastic process and suppose we have some model

$$E(X_n \mid \mathscr{F}_{n-1}) = f_n = f_n(\theta),$$

where $\theta \in \mathbb{R}^r$, and $f_n(\theta)$ is some $\mathscr{F}_{n-1}$ measureable function which is twice differentiable. Further, let

$$R_n = X_n - f_n(\theta)$$

and suppose that for $\sigma \in \mathbb{R}^s$,

$$E(R_n^2 \mid \mathscr{F}_{n-1}) = g_n^2 = g_n^2(\theta, \sigma).$$

In our applications we estimate the parameters θ by the conditional least squares estimators $\hat{\theta}_n$, which are solutions of the estimating equations

$$W_n(\theta) = \sum_{j=1}^{n} \frac{\mathrm{d}f_j(\theta)}{\mathrm{d}\theta} R_j = 0,$$

which usually do not involve an estimation of scale and then separately estimate a measure of scale, σ, from the estimated residuals

$$\hat{R}_j = X_j - f_j(\hat{\theta}_n).$$

Let χ_0 be an even function and define $h_n = E\{\chi_0(R_n/g_n(\theta, \sigma)) \mid \mathscr{F}_{n-1}\}$ where this conditional expectation is computed according to some model for the process. Typically for such models h_n will be free of θ and σ. We then estimate σ by solving

$$Q_n(\hat{\theta}_n, \sigma) = \sum_{j=1}^{n} \{\chi_0(\hat{R}_j/g_j(\hat{\theta}^n, \sigma)) - h_j\} \frac{\mathrm{d}g_j(\hat{\theta}_n, \sigma)}{\mathrm{d}\sigma} = 0 \tag{2.1}$$

and denote this estimate of σ by $\hat{\sigma}$.

Our main practical concern is the detection of outliers and non-symmetric error distributions so that our model for the process under the null hypothesis will at least specify that the conditional distributions of the R_n given $\mathscr{F}_{n-1}$ are symmetric about zero. Then for any odd function ψ,

$$E(\psi(R_n/g_n)|\mathscr{F}_{n-1}) = 0 \tag{2.2}$$

so that the process

$$S_n^{(1)}(\theta, \sigma) = \sum_{j=1}^n \psi(R_j/g_j) \tag{2.3}$$

will be a martingale under the null hypothesis. Note that $S_n^{(1)}$ is a natural generalisation of the sign test to other measures of location.

Theorem 2.1 shows that under regularity conditions $a_n^{-1/2} S_n^{(1)}(\hat{\theta}_n, \hat{\sigma}_n)$ is asymptotically normal with variance e and we use this asymptotic variance to construct an outlier detection test, typically under some model for R_n/g_n. That is if

$$|a_n^{-1/2} S_n^{(1)}(\hat{\theta}_n, \hat{\sigma}_n)| > z_{a/2} e^{\frac{1}{2}},$$

we would conclude that for some n the $f_n(\theta)$ are not solutions of $E(\psi(R_n)/g_n | \mathscr{F}_{n-1}) = 0$. This may then be due to the presence of asymmetric outliers or error distributions. Simulations in Section 4 below reveal that tests based on $S_n^{(1)}$ for typical ψ associated with measures of location can have poor power unless the estimating equations that define $\hat{\theta}_n$ include the constraint that $\sum R_j = 0$ and that some sensitivity is lost when the contamination is symmetric.

In order to increase the power and generality of our tests, we consider a statistic related to the robustified score statistic of Basawa et al. (1985). Let $\hat{\theta}_n$ and $\hat{\sigma}_n$ be as above and let

$$S_n^{(2)}(\theta, \sigma) = \sum_{j=1}^n w_j \psi(R_j/g_j)$$

for some vector of predictable weights $w_j \in \mathbb{R}^r$ be a martingale estimating equation. Typically, these estimating equations are constructed to guard against various types of outliers. Under the regularity conditions of Theorem 2.1, the statistic

$$a_n^{-1/2} S_n^{(2)}(\hat{\theta}_n, \hat{\sigma}_n)^{\mathrm{T}} e^{-1} S_n^{(2)}(\hat{\theta}_n, \hat{\sigma}_n) a_n^{-1/2}$$

will have asymptotically a chi-square distribution. Once again large values of this statistic lead us to conclude that the $f_n(\theta)$ are not solutions of $E(\psi(R_n | \mathscr{F}_{n-1}) = 0$. Note that if σ^2 were known then a Taylor series expansion of $S_n^{(2)}(\tilde{\theta}_n, \sigma^2)$, where $\tilde{\theta}_n$ is a solution of $S_n^{(2)}(\theta, \sigma^2) = 0$, about $\hat{\theta}_n$ shows that tests based on $S_n^{(2)}$ are asymptotically equivalent to testing if the means of the distributions of $\hat{\theta}_n$ and $\tilde{\theta}_n$ are the same, i.e. testing if $\hat{\theta}_n$ and $\tilde{\theta}_n$ are estimating the same parameters. The simulations of Section 4 show that for the examples considered the power of tests based on $S_n^{(2)}$ is superior to that of the sign test or that of tests based on $S_n^{(1)}$.

Our main technical result involves the joint distribution of a test statistic and location and scale parameters. We suppose throughout that a_n, I_n and $\tilde{I}_n$ are diagonal

matrices and further suppose that ψ' and χ_0' exist so that Taylor's theorem or the mean value theorem is applicable. A suitable ψ is Tukey's bisquare. For a function $g(\beta) \in \mathbb{R}^r$ where $\beta \in \mathbb{R}^s$ we take $dg(\beta)/d\beta$ to be the $r \times s$ matrix with (j, k) element $dg_j/d\beta_k$. The theorem is stated in terms of a general S_n which may be taken to be either of $S_n^{(1)}$ or $S_n^{(2)}$. In the former case note that S_n, a_n and w_n are all one dimensional.

Theorem 2.1. *If*

$$a_n^{-1/2} S_n(\theta, \sigma) \xrightarrow{d} N(0, \eta^2), \tag{2.4}$$

$$I_n^{-1/2} W_n(\theta, \sigma) \xrightarrow{d} N(0, \Sigma_1), \tag{2.5}$$

$$\tilde{I}_n^{-1/2} Q_n(\theta, \sigma) \xrightarrow{d} N(0, \Sigma_2) \tag{2.6}$$

$$a_n^{-1/2} \sum_{j=1}^n w_j \left(\frac{df_j}{d\theta}\right)^{\mathrm{T}} E(R_j \psi(R_j/g_j) \,|\, F_{j-1}) I_n^{-1/2} \to C_1, \tag{2.7}$$

$$I_n^{-1/2} \sum_{j=1}^n \frac{df_j}{d\theta} \left(\frac{dg_j}{d\sigma}\right)^{\mathrm{T}} E(R_j \chi_0(R_j/g_j) \,|\, F_{j-1}) \tilde{I}_n^{-1/2} \xrightarrow{P} C_2, \tag{2.8}$$

$$a_n^{-1/2} \sum_{j=1}^n w_j \frac{dg_j}{d\sigma}^{\mathrm{T}} E(\psi(R_j/g_j) \chi_0(R_j/g_j) \,|\, F_{j-1}) \tilde{I}_n^{-1/2} \xrightarrow{P} C_3, \tag{2.9}$$

$$L_{1n}(\hat{\theta}_n, \hat{\sigma}_n) = -a_n^{-1/2} \frac{dS_n}{d\theta}\bigg|_{\hat{\theta}_n, \hat{\sigma}_n} I_n^{-1/2} \xrightarrow{P} L_1(\theta, \sigma), \tag{2.10}$$

$$L_{2n}(\hat{\theta}_n, \hat{\sigma}_n)(\theta, \sigma) = a_n^{-1/2} \frac{dS_n}{d\sigma}\bigg|_{\hat{\theta}_n, \hat{\sigma}_n} I_n^{-1/2} \xrightarrow{P} L_2(\theta, \sigma), \tag{2.11}$$

$$G_{1n}(\hat{\theta}_n, \hat{\sigma}_n) = I_n^{-1/2} \frac{dW_n}{d\theta}\bigg|_{\hat{\theta}_n, \hat{\sigma}_n} I_n^{-1/2} \xrightarrow{P} G_1(\theta, \sigma), \tag{2.12}$$

$$G_{2n}(\hat{\theta}_n, \hat{\sigma}_n) = \tilde{I}_n^{-1/2} \frac{dQ_n}{d\sigma}\bigg|_{\hat{\theta}_n, \hat{\sigma}_n} \tilde{I}_n^{-1/2} \xrightarrow{P} G_2(\theta, \sigma), \tag{2.13}$$

$$G_{3n}(\hat{\theta}_n, \hat{\sigma}_n) = \tilde{I}_n^{-1/2} \frac{dQ_n}{d\theta} I_n^{-1/2} \xrightarrow{P} G_3(\theta, \sigma), \tag{2.14}$$

where G_1 and G_2 are non-singular. Let $B = (L_1 G_1^{-1} - L_2 G_2^{-1} G_3 G_1^{-1})$. Then

$$a_n^{-1/2} S_n(\hat{\theta}_n, \hat{\sigma}_n) \xrightarrow{d} N(0, e),$$

where

$$e = \eta^2 + B\Sigma_1 B^{\mathrm{T}} + L_2 G_2^{-1} \Sigma_2 G_2^{-\mathrm{T}} L_2^{\mathrm{T}} - C_1 B^{\mathrm{T}} - BC_1^{\mathrm{T}} - C_3 G_2^{-\mathrm{T}} L_2^{\mathrm{T}} - L_2 G_2^{-1} C_3^{\mathrm{T}}$$
$$+ BC_2 G_2^{-\mathrm{T}} L_2^{\mathrm{T}} + L_2 G_2^{-1} C_2^{\mathrm{T}} B^{\mathrm{T}}. \tag{2.15}$$

Proof. Taylor series expansions of $S_n(\hat{\theta}_n, \hat{\sigma}_n)$ around (θ, σ), $W_n(\hat{\theta}_n)$ about θ, $Q_n(\hat{\theta}_n, \hat{\sigma}_n)$ about (θ, σ) and (2.10)–(2.14) show that the asymptotic distribution of $a_n^{-1/2} S_n(\hat{\theta}_n, \hat{\sigma}_n)$ is the same as that of

$$a_n^{-1/2} S_n(\theta, \sigma) - B I_n^{-1/2} W_n - L_2 G_2^{-1} \tilde{I}_n^{-1/2} Q_n.$$

The theorem now follows from Cramér Wold device and the martingale central limit theorem using (2.4)–(2.9).

Further, note that for many ψ, under our hypothesis of symmetric errors, the asymptotic variance may be considerably simplified.

Corollary 2.2. *If the conditional distributions of the R_n given $\mathscr{F}_{n-1}$ are symmetric about zero and if ψ is an odd function then*

$$e = \eta^2 + B\Sigma_1 B^{\mathrm{T}} + L_2 G_2^{-1} \Sigma_2 G_2^{-\mathrm{T}} L_2^{\mathrm{T}} - C_1 B^{\mathrm{T}} - B C_1.$$

Remarks. (1) Note that under the conditions of Corollary 2.2, $L_{2n}(\theta, \sigma)$ is a martingale and the law of large numbers for martingales will often imply that $L_2(\theta, \sigma) = 0$. Similar results hold for $G_3(\theta, \sigma)$ as χ'_0, is an odd function. Thus one will often have that

$$e = \eta^2 + L_1 G_1^{-1} \Sigma_1 G_1^{-\mathrm{T}} L_1^{\mathrm{T}} - C_1 G_1^{-\mathrm{T}} L_1^{\mathrm{T}} - L_1 G_1^{-1} C_1.$$

In this case the asymptotic distribution of our test statistics will not depend on the distribution of $\hat{\sigma}_n$. This is true of all the examples we consider below and in such cases for practical convenience we estimate σ^2 by the median absolute deviations multiplied by $\Phi^{-1}(\frac{3}{4})$ which is still an M-estimator, see Hampel et al. (1986, p. 107).

(2) Conditions (2.10)–(2.13) generally require consistency of $\hat{\theta}_n$ and $\hat{\sigma}_n$ and a continuity condition on $L_1(\theta, \sigma)$, etc. See Klimko and Nelson (1978), Nelson (1980), Section 6.3 of Hall and Heyde (1980) and Crowder (1986) for related results.

3. Examples

Example 3.1. First-order autoregression, $\beta = 0.4$, $\sigma^2 = 1$.

We examine a first order autoregression $X_n = \beta X_{n-1} + \varepsilon_n$ where $|\beta| < 1$, and the ε_n are independently and identically distributed as standard normal variables with mean 0 and variance σ^2.

Let $\hat{R}_j = X_j - \hat{\beta} X_{j-1}$. We consider the statistics

$$T_0 = n^{-1/2} \sum_{j=1}^{n} sign(R_j),$$

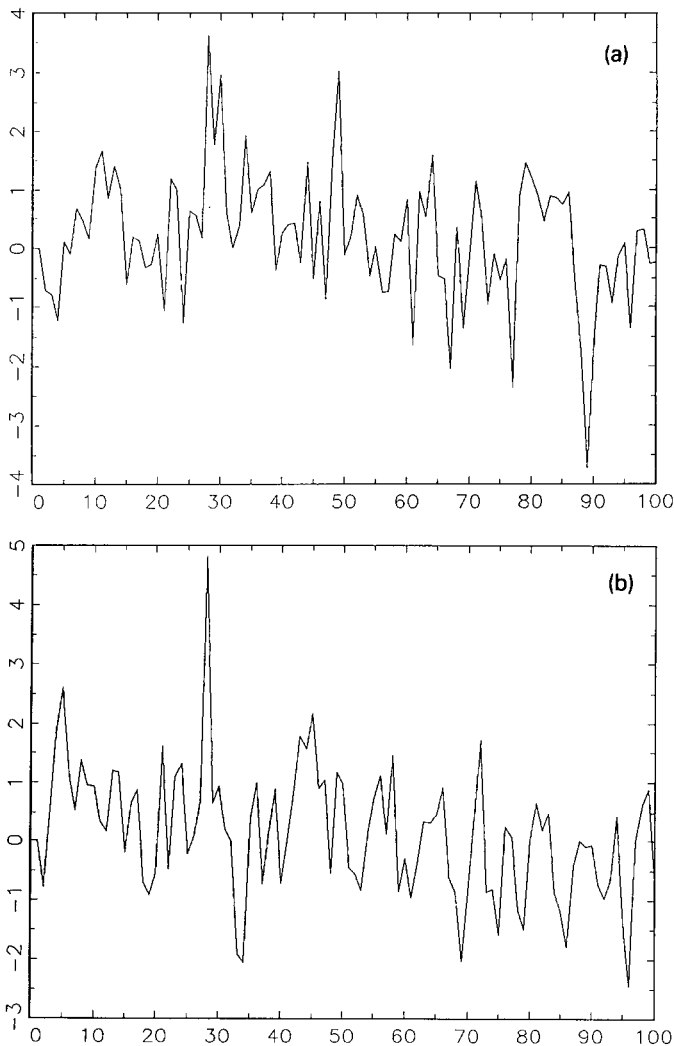$$T_1 = n^{-1/2} S_n^{(1)} = n^{-1/2} \sum_{j=1}^{n} \psi\left(\frac{\hat{R}_j}{\hat{\sigma}}\right)$$

Fig. 1. Outcomes of a first-order autoregressive process: (a) outlier detected; (b) outlier not detected.

and

$$T_2 = n^{-1/2} S_n^{(2)} = n^{-1/2} \sum_{j=1}^{n} X_{j-1} \psi\left(\frac{\hat{R}_j}{\hat{\sigma}}\right).$$

It is easily shown that the asymptotic distributions of T_0, T_1, and T_2 are normal with zero mean and variances 1, $E(\psi(Z)^2)$, and $V_1 \sigma^2/(1 - \beta^2)$, where

$$V_1 = [E(\psi^2(Z)) + E^2(\psi'(Z) - 2E(\psi'(Z))E(Z\psi(Z))], \tag{3.1}$$

respectively where Z has a standard normal distribution. Note that in this example for T_1 both C_1 and L_1 are zero so that S_n and W_n are uncorrelated. This leads to the expectation, confirmed in the simulations below, that tests based on T_0 and T_1 will be useless in this case.

Two simulated outcomes of this process with one additive outlier of size $+4$ at point 28 is given in Fig. 1. For the data of Fig. 1(a) the value of T_2 was -0.6774 which was less than $-1.96 \times V_1 \hat{\sigma}^2 / (1 - \hat{\beta}^2) = -0.426$. In this case the least squares estimate of β was 0.3486 whilst a robust estimate using Huber's ψ with $k = 1.7$ and the median absolute deviation multiplied by $\Phi^{-1}(\frac{3}{4})$ to estimate scale was 0.2987. In Fig. 1(b) the value of T_2 was -0.365 and $-1.96 \times V_1 \hat{\sigma}^2 / (1 - \hat{\beta}^2) = 0.472$. Here the least squares estimate was 0.3969 and a robust estimator computed using Hubers ψ as above was 0.3788. These simulated outcomes illustrate how the procedure only detects the presence of outliers that are influencing the parameter estimates rather than identifying particular outliers.

Example 3.2. First-order autoregression, $\beta = 0.4$, $\sigma = 1$ with mean $\alpha = 1$.

We consider the application of our results to the first order autoregressive process $X_n = \alpha + \beta X_{n-1} + R_n$ where $-\infty < \alpha < \infty$, $|\beta| < 1$ and the R_n form a sequence of independently and identically distributed random variables with zero means and common variance σ^2. In this example the estimating equations for $\hat{\theta}_n$ include the constraint that $\sum R_j / g_j = 0$. We retain the outlier construction of Example 3.1.

Now $\hat{R}_j = X_j - \hat{\alpha}_n - \hat{\beta} X_{j-1}$ and we let

$$T_0 = n^{-1/2} \sum_{j=1}^{n} sign(\hat{R}_j),$$

$$T_1 = n^{-1/2} S_n^{(1)} = n^{-1/2} \sum_{j=1}^{n} \psi \left(\frac{\hat{R}_j}{\hat{\sigma}} \right)$$

and

$$T_2 = n^{-1/2} S_n^{(2)} = n^{-1/2} \sum_{j=1}^{n} \begin{pmatrix} 1 \\ X_{j-1} \end{pmatrix} \psi \left(\frac{\hat{R}_j}{\hat{\sigma}} \right).$$

The asymptotic variance of T_0 was shown in Huggins (1989) to be $1 - 2/\pi$ and that of T_1 can be shown to be V_1, given by (3.1). The asymptotic distribution of T_2 is bivariate normal with covariance matrix

$$V_1 \begin{pmatrix} 1 & \alpha/(1 - \beta) \\ \alpha/(1 - \beta) & \alpha^2/(1 - \beta)^2 + \sigma^2/(1 - \beta^2) \end{pmatrix},$$

where again V_1 is given by (3.1).

Example 3.3. A regression model with autoregressive errors.

In this example we illustrate how our results may be applied in more complex situations. The works of Basawa et al. (1985) and Kulkarni and Heyde (1987) have both considered the regression model with autoregressive errors,

$$X_n = \beta^T C_n + Y_n, \qquad Y_n = \alpha Y_{n-1} + \varepsilon_n,$$

where $|\alpha| < 1$ and the ε_n are independently $N(0, 1)$ random variables. We consider here the simple regression

$$X_n = \beta_1 + \beta_2 C_n + Y_n$$

and take $a_n = I_n = n$.

Then

$$W_n = \sum_{j=1}^{n} \begin{pmatrix} 1 - \alpha \\ C_j - \alpha C_{j-1} \\ Y_{j-1} \end{pmatrix} \varepsilon_j.$$

Following Basawa et al. (1985) and Kulkarni and Heyde (1987) let

$$\tilde{\varepsilon}_n = \psi(\varepsilon_n), \qquad \tilde{Y}_n = \sum_{j=1}^{n-1} \alpha^j \tilde{\varepsilon}_{n-j},$$

where $\varepsilon_n = X_n - \alpha X_{n-1} - \beta_1(1 - \alpha) - \beta_2(C_n - \alpha C_{n-1})$, and let

$$S_n = \sum_{j=1}^{n} \begin{pmatrix} 1 - \alpha \\ C_j - \alpha C_{j-1} \\ \tilde{Y}_{j-1} \end{pmatrix} \tilde{\varepsilon}_j.$$

We suppose that $c = \lim\{n^{-1}\sum(C_j - \alpha C_{j-1}\}$, $d = \lim\{n^{-1}\sum(C_j - \alpha C_{j-1}\}$ and $\lim n^{-1}\sum C_j^2$ exist, and let $k_1 = E(Z\psi(Z))$, $k_2 = E(\psi^2(Z))$, where $Z \sim N(0, 1)$. Then it can be verified that

$$\eta^2 = k_2 \begin{bmatrix} (1 - \alpha)^2 & c & 0 \\ c & d & 0 \\ 0 & 0 & k_2/(1 - \alpha^2) \end{bmatrix}, \qquad \Sigma_1 = \begin{bmatrix} (1 - \alpha)^2 & c & o \\ c & d & 0 \\ 0 & 0 & 1/(1 - \alpha^2) \end{bmatrix},$$

$$L_1 = E(\psi'(Z)) \begin{bmatrix} (1 - \alpha)^2 & c & 0 \\ c & d & 0 \\ 0 & 0 & k_1/(1 - \alpha^2) \end{bmatrix},$$

$G_1 = \Sigma_1$ and $C_1 = L_1 k_1 / E(\psi'(Z))$.

4. Simulations

We consider here a small simulation study of Examples 3.1 and 3.2 to examine the powers of the tests using two types of contamination, additive and innovations outlier models, commonly used in other studies, Denby and Martin (1979), Martin and Yohai (1986), and Bruce and Martin (1989). An additive outlier is an outlier added to the process, that is if X_n is the stochastic process of interest we observe $Y_n = X_n + Z_n \xi_n$ where Z_n takes the values 0 or 1 and ξ_n is some contaminating process. For innovations outliers the model is $X_n = f_n(\theta) + \varepsilon_n + Z_n \xi_n$ where Z_n and ξ_n are as

above. In view of the differentiability requirements of Theorem 2.1 we use Tukey's bisquare

$$\psi(x) = \begin{cases} x(1 - (x/a)^2)^2 & \text{if } |x| < a, \\ 0, & \text{otherwise} \end{cases}$$

to illustrate our technique and the median absolute deviation multiplied by $\Phi^{-1}(\frac{3}{4})$ to estimate scale.

We consider outlier configurations consisting of a single additive outlier of $+4$ at point 28, a single innovations outlier of $+4$ at point 28, and an additive outlier of $+4$ at point 28 and another additive outlier of -4 at point 72. The results are based on 1000 simulations.

For Example 3.1 the tests were conducted at a nominal level of 0.1 and the null distribution of all three statistics gave approximately this significance. The power of T_0 and T_1 remained close to 0.1 for all three outlier configurations. The power of T_2 was approximately 0.34, for one additive outlier, approximately 0.3 for one innovations outlier, and approximately 0.4 for the two outliers of opposite sign. This example illustrates the clear superiority of T_2 without the constraint that $\sum R_j/g_j = 0$.

For Example 3.2 the simulated tests were again conducted at a nominal level of 0.1 and under the null distribution all tests had approximately this power. For one additive outlier the powers of T_0, T_1 and T_2 were approximately 0.12, 0.33 and 0.43, respectively, and were similar for the innovations outlier. In the symmetric outliers case the powers were approximately 0.1, 0.17 and 0.42, respectively. If the contamination consisted of three additive outliers of size $+4$ at points 28, 50 and 72 the power of the tests were approximately 0.26, 0.87 and 0.88. If the additive outlier at point 50 was changed to -4 the powers were 0.11, 0.34 and 0.54.

Acknowledgement

References

I.V. Basawa, Neyman–Le Cam tests based on estimating functions, in: L. Le Cam and R. Olshen, eds., Proc. of Berkeley Conference in Honor of J. Neyman and J. Kiefer, Vol. 1 (Wadsworth, Belmont, CA, 1985).

I.V. Basawa, R.M. Huggins and R.G. Staudte, Robust tests for time series with an application to first-order autoregressive processes, Biometrika 72 (1985) 559–571.

B.M. Brown, A short cut test for outliers using residuals, Biometrika 62 (1975) 623–629.

B.M. Brown and D.G. Kildea, Outlier-detection tests and robust estimators based on signs of residuals, Comm. Statist. Theory Methods A8(3) (1979) 257–269.

A.G. Bruce and R.D. Martin, Leave-k-out diagnostics for time series (with discussion), J. Roy. Statist. Soc. B 53 (1989) 363 424.

O.H. Bustos, General M-estimates for contaminated pth-order autoregressive process: Consistency and asymptotic normality, Z. Wahrs. V. Geb. 59 (1982) 491–504.

M.J. Crowder, On consistency and inconsistency of estimating equations, Econ. Theory 2 (1986) 305–330.

H.T. David, The sample mean among the moderate order statistics, Ann. Math. Statist. 33 (1962) 1160–1166.

L. Denby and R.D. Martin, Robust estimation of the first order autoregressive parameter, J. Amer. Statist. Soc. 74 (1979) 140–146.

A.J. Fox, Outliers in time series, J. Roy Statist. Soc. B, 34 (1972) 350–363.

V.P. Godambe, The foundations of finite sample estimation for discrete stochastic processes, Biometrika 72 (1985) 419–428.

V.P. Godambe and C.C. Heyde, Quasi-likelihood and optimal estimation, Internat. Statist. Rev. 55 (1987) 231–244

P. Hall and C.C. Heyde, Martingale Limit Theory and its Application (Academic Press, London, (1980).

F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, Robust Statistics (Wiley, New York, 1986).

R.M. Huggins The sign test for stochastic processes, Aust. J. Statist. 31 (1989) 153–165.

L.A. Klimko and P.I. Nelson, On conditional least squares estimation for stochastic processes, Ann. Statist. 6 (1978) 629–642.

P.M. Kulkarni and C.C. Heyde, Optimal robust estimation for discrete time stochastic processes, Stoch. Proc. App. 26 (1987) 267–276.

R.D. Martin, Robust estimation of autoregressive models (with discussion), in: D.R. Brillinger and G.C. Tiao, eds., Directions in Time Series (Inst. Math. Statist., Haywood, 1980) pp. 228–262.

R.D. Martin, The Cramér–Rao bound and robust M-estimates for autoregressions, Biometrika 69 (1982) 437–442.

R.D. Martin and V.J. Yohai, Robustness in time series and estimating ARMA models, in: E.J. Hannan, P.R. Krishnaiah and M.M. Rao, eds., Handbook of Statistics (Elsevier New York, 5th ed., 1985) pp. 119–155.

R.D. Martin and V.J. Yohai, Influence functionals for time series (with discussion), Ann. Statist. 14 (1986) 781–855.

P.I. Nelson, A note on the strong consistency of least squares estimators in regression models with martingale difference errors, Ann. Statist. 8 (1980) 1057–1064.