International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 3 (Special Issue), 2004, pp 23–38

The Squared Residuals Process

Ian B. MacNeill
Department of Statistical and Actuarial Sciences
University of Western Ontario
London, Canada, N6A 5B9

[Received April 22, 2004; Accepted August 4, 2004]

Abstract

The partial sums of squared regression residuals are used to define the squared residuals process. This process is characterized in terms of regressor functions, the serial correlation structure and the distribution of the noise process. It is shown that these factors affect the process independently thus making it possible to investigate separately the effects of regression, serial correlation and non-normality, and then to combine them as required to determine joint effects. Since the total sum of squared regression residuals is a particular partial sum, the results of this paper apply to a variety of second moment statistics such as variance estimators in regression problems. The results concerning the effects of serial correlation and/or non-normality on such statistics expand upon previously available results. The squared residuals process finds application in the change-point problem.

Keywords and Phrases: Regression residuals, Squared regression residuals, Cumulative sums, Partial sums, Serial correlation, Non-normality, Cumulants, Stationary time series, Polynomial regression, Harmonic regression, Heteroscedasticity.

AMS Classification: Primary 62J05; Secondary 62M10.

1 Introduction

This paper examines the properties of partial sums, or cusums, of squares of regression residuals. These partial sums are used to define a stochastic process referred to below as the squared residuals process. This process is characterized in terms of: the regressor functions; the serial correlation structure of the noise process; and the distribution of the noise process which, for linear time series, is characterized by the fourth cumulant of the innovations series.

Our results indicate these three factors affect the squared residuals process independently; hence one can study separately the large sample effects of regression, serial correlation and non-normality, and then combine them as required to determine joint effects. The linear regression models we consider are defined in terms of non-stochastic regressor functions of time. The family of models considered is sufficiently large that it includes all commonly used models of the type. The noise models considered are stationary time series satisfying a cumulant condition; specific analysis is provided for linear time series, which include ARMA models.

Examples of the use of the methodology are presented for several common regression models where, inter alia, the effects of serial correlation and/or non-normality in the noise process are examined.

Since the total sum of squared regression residuals is a particular partial sum, the results presented below apply to a variety of second moment statistics such as variance estimators in regression problems. Our results concerning the effects of regression, serial correlation and non-normality on such statistics expand upon previously available results. Since these effects can be substantial they should not be ignored.

The sequence of partial sums of (unsquared) regression residuals and the corresponding residual processes have been treated by MacNeill (1978a,b) and Jandhyala and MacNeill (1989). A principal application area for residual processes is the change-point problem. These statistics are usually defined in terms of partial sums of regression residuals, and are derived using the Bayes-type methods introduced to this area by Chernoff and Zacks (1964). Hsu (1977) applied these methods to derive a test for variance shift in a series of independent observations; the test is based on the reversed sequence of partial sums of squared residuals. Inclán and Tiao (1994) studied the problem of multiple change points in the variance of a sequence of independent observations; their tests are based on iterative cumulative sums of squares. Tsay (1988) discussed variance changes for autoregressive-moving average models. A closely related problem is that of testing for heteroscedasticity. An example of such tests, which are usually based on ordinary least squares residuals, is that of Harrison and McCabe (1979). A related problem involving partial sums of lagged cross-products of regression residuals is discussed by De Gooijer and MacNeill (1999).

A substantial literature exists on the properties and applications of partial sums of residuals, but although the total sum of squares of residuals is ubiquitous in statistics, the extent of the literature on partial sums of squared residuals is quite modest. The joint effects of non-normality and general serial correlation on sums of squares of regression residuals seem not to have been examined. These effects are studied and discussed in the context of tests for variance shifts at unknown times and for heteroscadesticity

The plan of this paper is as follows. In section 2 we present the regression models to be considered, and define the "squared residuals process" and two related processes.

The large sample properties of these processes are investigated in sections 3, 4, and 5. In section 6 we discuss the properties of cumulant expressions which arise in our discussion of the joint effects of non-normality and serial correlation. Variable sampling rates are discussed in section 7. In section 8 we apply our results to commonly used linear regression models. Sections 9 contains a discussion of the use of the squared residual processes to the change-point problem.

2 Regression Models

The error structure we consider for linear regression models is that provided by a zero-mean, discrete time, stationary time series, $X(j)$ $(j = 0, \pm 1, \pm 2, \ldots)$. Results derived for this error structure, as opposed to independent, identically distributed, normal random variables, allow one to examine the large sample effects on regression of serial correlation and non-normality. Of course, results for the standard normal regression model are obtained as a special case. If we let $\{g_k(t), t \in [0,1]\}$ $(k = 0, \ldots, p)$ be a collection of non-stochastic regressor functions of time, we can define the triangular array $Y_n(j)$ $(n > p,\ j = 1, \ldots, n)$ of dependent variables as follows:

$$Y_n(j) = \sum_{i=0}^{p} \beta_i g_i(j/n) + X(j).$$

To begin, the total time of observation has been compressed to the interval [0,1], and observations are assumed to have been taken at equi-spaced time points; variable sampling rates are discussed in section 7. The matrix formulation of this model is:

$$\boldsymbol{Y}_n = \boldsymbol{A}_n \boldsymbol{\beta}_p + \boldsymbol{X}_n \tag{1}$$

where the (j, i)th component of the design matrix is $g_i(j/n)$. The regression parameter estimators are denoted by:

$$\tilde{\boldsymbol{\beta}}_{pn} = (\boldsymbol{A}_n' \boldsymbol{A}_n)^{-1} \boldsymbol{A}_n' \boldsymbol{Y}_n.$$

We denote sequences of partial sums of squared regression residuals by $R_{gX_n}(j)$ $(1 \leq j \leq n,\ n \geq 1)$ where

$$\begin{aligned} R_{gX_n}(j) &= \sum_{i=1}^{j} \left\{ Y_n(i) - \tilde{Y}_n(i) \right\}^2 = \sum_{i=1}^{j} r_{gX_n}^2(i) \quad , \\ \hat{Y}_n(i) &= \tilde{\boldsymbol{\beta}}_{pn}' \boldsymbol{g}(i/n) \end{aligned}$$

and

$$\boldsymbol{g}'(i/n) = \{g_0(i/n), g_1(i/n), \ldots, g_p(i/n)\}.$$

Under the mild restriction that the Riemann integrals on [0,1] of $g_k^2(\cdot)$ exist, then the (r_1, r_2)th component of $\lim_{n\to\infty} n^{-1}(\boldsymbol{A}_n' \boldsymbol{A}_n) \equiv \boldsymbol{G}$ is $\int_0^1 g_{r_1}(t) g_{r_2}(t)\, dt$. If in addition the regressor functions are linearly independent then the inverse of $\boldsymbol{G}$ exists. Under these circumstances we define a bilinear form $g(s,t)$ as follows:

$$g(s,t) = \boldsymbol{g}'(s)\boldsymbol{G}^{-1}\boldsymbol{g}(t).$$

Since we shall be concerned with weak convergence in the space of functions continuous on the unit interval, $C[0,1]$, we use these sequences of partial sums of squared residuals to define a sequence of stochastic processes $\{\rho_{gX_n}(t),\ t \in [0,1],\ n \geq 1\}$ possessing continuous sample paths as follows:

$$\sqrt{n}\rho_{gX_n}(t) = R_{gX_n}([nt]) + (nt - [nt])^2 r^2_{gX_n}([nt]+1) \tag{2}$$

where $[nt]$ is the integer part of nt. The processes defined by (2) will be referred to as "squared residuals processes" and will be a principal focus of our interest.

If $\boldsymbol{I}_n(t)$ is an $n \times n$ matrix with the first $[nt]$ diagonal elements equal to 1, the next equal to $nt - [nt]$ and all other elements equal to zero, then

$$\sqrt{n}\rho_{gX_n}(t) = \boldsymbol{X}_n' \left\{\boldsymbol{I}_n - \boldsymbol{A}_n(\boldsymbol{A}_n'\boldsymbol{A}_n)^{-1}\boldsymbol{A}_n'\right\} \boldsymbol{I}_n(t)\boldsymbol{I}_n(t) \left\{\boldsymbol{I}_n - \boldsymbol{A}_n(\boldsymbol{A}_n'\boldsymbol{A})^{-1}\boldsymbol{A}_n'\right\} \boldsymbol{X}_n.$$

Where there is no confusion we will simplify the notation by dropping the subscripts on the regression matrices. Of course the expression above is not changed by replacing $\boldsymbol{X}_n$ with $\boldsymbol{Y}_n$.

We define a related sequence of stochastic processes $\{\chi^2_{X_n}(t), t \in [0,1]\}$ $(n = 1,2,\ldots)$ as follows:

$$\sqrt{n}\chi^2_{X_n}(t) = \sum_{j=1}^{[nt]} X^2(j) + (nt - [nt])^2 X^2([nt]+1). \tag{3}$$

The process defined by (3) will be referred to as the "squared errors process".

Another related sequence of stochastic processes is defined as follows:

$$\begin{aligned} D_{gX_n}(t) \quad = \quad & \boldsymbol{X}_n'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{I}(t)\boldsymbol{I}(t)\boldsymbol{X}_n + \boldsymbol{X}_n'\boldsymbol{I}(t)\boldsymbol{I}(t)\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}_n \\ & -\boldsymbol{X}_n'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{I}(t)\boldsymbol{I}(t)\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}_n. \end{aligned} \tag{4}$$

The process $\{D_{gX_n}(t),\ t \in (0,1]\}$ characterizes the effects of regression on the squared residuals process; we refer to this process as the "squared regression effects process". We then have

$$\sqrt{n}\rho_{gX_n}(t) = \sqrt{n}\chi^2_{X_n}(t) - D_{gX_n}(t). \tag{5}$$

We return to discussion of the noise process $X(n)$ $(n = 0, \pm 1, \ldots)$. The covariance function for this series is $R_X(\nu) = E\{X(n)X(n+\nu)\}$, $|\nu| < \infty$. If the covariance function is absolutely summable, then the spectral density function $f_X(\lambda) =$

$\frac{1}{2\pi}\sum_{|\nu|<\infty} e^{-\lambda\nu}R_X(\nu)$, $\lambda \in [-\pi, \pi]$, exists. If the spectral density is positive, that is, if

$$f_X(\lambda) \geq a > 0 \quad , \quad \lambda \in [-\pi, \pi], \tag{6}$$

then the process can be expressed either as an infinite moving average or as an infinite autoregression; that is it is invertible.

In the sequel we require a central limit theorem for time series. Conditions that guarantee convergence in distribution of $n^{-1/2}\sum_{j=1}^{[nt]} X(j),\ t \in [0,1]$, to the normal with zero mean and variance $\{2\pi f_X(0)t\}$, are those given by Brillinger (1973). These conditions are stated in terms of cumulant functions which are defined as follows:

$$C_{k+1}(\nu_1, \ldots, \nu_k) = \text{Cum}\{X(n+\nu_1), X(n+\nu_2), \ldots, X(n+\nu_k), X(n)\}. \tag{7}$$

Stationarity to order $k+1$ is implicit in this definition. When necessary we assume the cumulants exist and satisfy what we call the Brillinger conditions, namely,

$$|C_{k+1}(\nu_1, \nu_2 \ldots, \nu_k)| < \frac{L_k}{\prod_{j=1}^{k}(1+\nu_j^2)} \tag{8}$$

for some finite L_k, $k = 1, 2, \ldots$.

We let $S_{X_n}(j) = \sum_{i=1}^{j} X(i)$ and define a sequence of stochastic processes $\{\theta_{X_n}(t), t \in [0,1]\}\{n = 1, \ldots\}$ possessing continuous sample paths by

$$n^{\frac{1}{2}}\theta_{X_n}(t) = S_{X_n}([nt]) + (nt - [nt])X([nt]+1). \tag{9}$$

If a stochastic process $\{B_X(t),\ t \in [0,1]\}$ is defined by $B_X(t) = \{2\pi f_X(0)\}^{\frac{1}{2}}B(t)$ where $B(\cdot)$ is standard Brownian motion then under conditions (6) and (8) Tang and MacNeill (1993) have shown that $\{\theta_{X_n}(t),\ t \in [0,1]\}$ $(n = 1, 2, \ldots)$ converges weakly to $\{B_X(t),\ t \in [0,1]\}$; that is

$$\theta_{X_n}(t) \Rightarrow B_X(t) = \{2\pi f(0)\}^{\frac{1}{2}}B(t).$$

The measure on $C[0,1]$ corresponding to standard Brownian motion is Wiener measure denoted by W. We denote the measure corresponding to $B_X(\cdot)$ by W_X.

3 The Squared Errors Process

Since the process $\{\chi^2_{X_n}(t),\ t \in [0,1]\}$ has a non-zero mean value function that grows with increasing sample size we define a related zero-mean process $\{\tilde{\chi}^2_{X_n}(t),\ t \in [0,1]\}$ as follows:

$$\tilde{\chi}^2_{X_n}(t) = \chi^2_{X_n}(t) - E\left\{\chi^2_{X_n}(t)\right\} \ ,$$

where

$$E\left\{\chi^2_{X_n}(t)\right\} = \sqrt{n}tR_X(0) + O(1/n).$$

If $nt = [nt]$ then the approximation is exact. We let $[nt] = j$ and $[ns] = k$. If the Brillinger condition (8) is assumed to hold then the covariance kernel for the process may be approximated as follows:

$$\begin{aligned} K_{X_n}(s,t) &\equiv E\left\{\tilde{\chi}^2_{X_n}(t)\tilde{\chi}^2_{X_n}(s)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{j}\sum_{l=1}^{k} E\left\{X^2(i)X^2(l)\right\} - \frac{1}{n}jkR^2_X(0) + O(1/n) \\ &= \frac{2}{n}\sum_{i=1}^{j}\sum_{l=1}^{k} R^2_X(i-l) + \frac{1}{n}\sum_{i=1}^{j}\sum_{l=1}^{k} \text{Cum}\ \{X(i)X(i)X(l)X(l)\} + O(1/n). \end{aligned}$$

Again, if $nt = [nt]$ and $ns = [ns]$ then the approximation is exact. Furthermore, with the notation $s \wedge t \equiv \min(s,t)$, condition (8) implies that:

$$\frac{2}{n}\sum_{i=1}^{j}\sum_{l=1}^{k} R^2_X(i-l) = 2(j/n \wedge k/n)\sum_{|\nu|<\infty} R^2_X(\nu) + O(1/n)$$

and

$$\sum_{i=1}^{j}\sum_{l=1}^{k} \text{Cum}\ \{X(i)X(i)X(l)X(l)\} = (j/n \wedge k/n)\sum_{|\nu|<\infty} C_{X4}(0,\nu,\nu) + O(1/n)\ .$$

We can now observe that, uniformly in s and t,

$$K_{X_n}(s,t) \longrightarrow (s \wedge t)K_X$$

where

$$K_X = 4\pi\int_{-\pi}^{\pi} f^2_X(\lambda)\, d\lambda + F_{X1} \tag{10}$$

and where

$$F_{X1} = \sum_{|\nu|<\infty} C_{X4}(0,\nu,\nu). \tag{11}$$

We establish asymptotic normality in the following theorem.

Theorem 1: *Under assumptions (6) and (8) the k-vector* $\{\tilde{\chi}^2_{X_n}(t_1), \ldots, \tilde{\chi}^2_{X_n}(t_k)\}$ *has a non-trivial asymptotic probability distribution that is normal with zero mean and covariance matrix* $K_X \parallel t_i \wedge t_j \parallel$.

Proof: The covariance matrix is derived directly from (10). The Brillinger (1973) condition (8) can be used to demonstrate that the cumulants of orders higher than two of a vector component of $\tilde{\chi}^2_{X_n}(t_i)$ are $O(n^{-1/2})$ or smaller and hence that $\tilde{\chi}^2_{X_n}(t_i)$ converges in distribution to the normal with zero mean and variance given by (10). The Cramér-Wold device for demonstrating asymptotic multivariate normality can be used to complete the proof for this k-dimensional case.

Tightness of the sequence of measures $P_{X^2_n}$ $(n = 1, 2, \ldots)$ generated in $C[0,1]$ by $\{\tilde{\chi}^2_n(t),\ t \in [0,1]\}$ $(n = 1, 2, \ldots)$ can be demonstrated by showing the existence of a constant $C_X > 0$ such that for any $t_2 > t_1$,

$$E\left\{\tilde{\chi}^2_n(t_2) - \tilde{\chi}^2_n(t_1)\right\}^4 \leq C_X(t_2 - t_1)^2. \tag{12}$$

This can be shown using arguments similar to those used to derive the covariance kernel. If the process $\{B_{X^2}(t),\ t \in [0,1]\}$ is defined by $B_{X^2}(t) = \{K_X\}^{1/2}B(t)$ and if W_{X^2} is the measure in $C[0,1]$ corresponding to $B_{X^2}(\cdot)$, then we have the following result.

Theorem 2: *Under assumptions (6) and (8)*

$$P_{X^2_n} \Longrightarrow W_{X^2}.$$

Proof: Theorem 1 assures us that the finite dimensional distributions of $P_{X^2_n}$ converge to those of W_{X^2}, and (12) implies that the sequence $P_{X^2_n}$ $(n = 1, 2, \ldots)$ is tight. The proof is completed by applying Theorem 12.3 of Billingsley (1968).

4 Squared Regression Effects Process

We now consider the large sample properties of $\{D_{gX_n}(t),\ t \in [0,1]\}$ as defined by (4). First we define a limit process $\{D_{gX}(t),\ t \in [0,1]\}$ as follows:

$$\begin{aligned} D_{gX}(t) = H_g\{B_X(t)\} &= 2\int_0^t \int_0^1 g(x,y)\, dB_X(x)\, dB_X(y) \\ &\quad - \int_0^t \int_0^1 \int_0^1 g(y,s)g(s,x)\, dB_X(x)\, dB_X(y)\, ds. \end{aligned} \tag{13}$$

Since $B_X(t) = \{2\pi f_X(0)\}^{1/2}B(t)$ we may write $D_{gX}(t) = \{2\pi f_X(0)\}H_g\{B(t)\}$. With some mild restrictions on $g(x,y)$, which we discuss below, the function $H_g(\cdot)$ as defined by (13) is a continuous function from $C[0,1]$ into itself. If P_{gX} denotes the measure generated in $C[0,1]$ by $H_g\{B(t)\}$ then in the notation of Theorem 5.5 of Billingsley (1968), $P_{gX} = W(H_g)^{-1}$ where W is Wiener measure. Also, using the functional form of (4), we define a sequence of processes $\{D_{gX_n}(t),\ t \in [0,1]\}$ $(n = 1, 2, \ldots)$

and sequences of continuous functions $H_{gX_n}(\cdot)$, $H_{gX_n1}(\cdot)$, $H_{gX_n2}(\cdot)$, and $H_{gX_n3}(\cdot)$, $(n = 1, 2, \ldots)$ from $C[0,1]$ into itself by

$$\begin{aligned} D_{gX_n}(t) &= H_{gX_n}\{\theta_{X_n}(t)\} \\ &= H_{gX_n1}\{\theta_{X_n}(t)\} + H_{gX_n2}\{\theta_{X_n}(t)\} + H_{gX_n3}\{\theta_{X_n}(t)\} \\ &= \boldsymbol{X}_n'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{I}(t)\boldsymbol{I}(t)\boldsymbol{X}_n + \boldsymbol{X}_n'\boldsymbol{I}(t)\boldsymbol{I}(t)\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}_n \\ &\quad -\boldsymbol{X}_n'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{I}(t)\boldsymbol{I}(t)\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}_n. \end{aligned}$$

If P_{gX_n} $(n = 1, 2, \cdots)$ denotes the set of measures generated in $C[0,1]$ by $H_{gX_n}\{\theta_{X_n}(t)\}$ then in the notation of Theorem 5.5 of Billingsley (1968), $P_{gX_n} = P_{X_n}(H_{gX_n})^{-1}$.

We now state the theorem that characterizes the large sample properties of the squared regression effects process.

Theorem 3: *Assume conditions (6) and (8). Further assume* $g_r(t)$ $(r = 0, 1, \ldots, p)$ *are linearly independent non-stochastic regressor functions that are continuously differentiable on* $[0,1]$. *Then:*

(i) $P_{gX_n} \Rightarrow P_{gX}$;

(ii) $E\{D_{gX_n}(t)\} \longrightarrow \{2\pi f_X(0)\}\int_0^t g(x,x)\,dx$; *and*

(iii) $Cov\{D_{gX_n}(t), D_{gX_n}(s)\} \longrightarrow \{2\pi f_X(0)\}^2 \Gamma_g(s,t)$

where

$$\Gamma_g(s,t) = 4\int_0^{s\wedge t)} g(x,x)\,dx - 2\int_0^t\int_0^s g^2(x,y)\,dx\,dy. \tag{14}$$

Proof: We note that

$$\frac{1}{\sqrt{n}}\boldsymbol{X}_n'\boldsymbol{A}_n \xrightarrow{D} \int_0^1 \boldsymbol{g}'(x)\,dB_X(x),$$

$$\frac{1}{\sqrt{n}}\boldsymbol{A}_n'\boldsymbol{I}_n(t)\boldsymbol{I}_n(t)\boldsymbol{X}_n \xrightarrow{D} \int_0^t \boldsymbol{g}'(y)\,dB_X(y),$$

and

$$n(\boldsymbol{A}_n'\boldsymbol{A}_n)^{-1} \longrightarrow \boldsymbol{G}^{-1}.$$

Hence it can be shown that $H_{gX_n}(\cdot)$ converges uniformly to $H_{gX}(\cdot)$. Proof of weak convergence follows by application of Theorem 5.5 of Billingsley (1968).

Since

$$E\{dB(x)\,dB(y)\} = \begin{cases} dx & \text{x=y} \\ 0 & \textit{otherwise}, \end{cases}$$

we observe that

$$E\left[D_{gX}(t)\right] = \{2\pi f_X(0)\} \int_0^t g(x,x)\,dx \quad .$$

To compute the covariance kernel we first note that

$$\begin{aligned} E&\left\{\int_0^t \int_0^1 g(x,y)\,dB(x)\,dB(y)\right\}\left\{\int_0^s \int_0^1 g(u,\nu)\,dB(u)\,dB(\nu)\right\} \\ &= \int_0^t \int_0^1 \int_0^s \int_0^1 g(x,y)g(u,\nu)\big[E\{dB(u)\,dB(\nu)\}E\{dB(x)\,dB(y)\} \\ &\qquad\qquad + E\{dB(u)\,dB(x)\}E\{dB(\nu)\,dB(y)\} \\ &\qquad\qquad + E\{dB(u)\,dB(y)\}E\{dB(x)\,dB(\nu)\}\big] \\ &= \int_0^t g(u,u)\,du \int_0^s g(x,x)\,dx + \int_0^{s\wedge t} g(x,x)\,dx + \int_0^t \int_0^s g^2(x,y)\,dx\,dy. \end{aligned}$$

We have made use of the fact that $\int_0^1 g^2(x,y)\,dy = g(x,x)$. Computations similar to the above applied to $E\{D_{gX}(t)D_{gX}(s)\}$ where $D_{gX}(\cdot)$ is defined by (13) will complete the proof of the theorem by verifying that

$$Cov\,\{D_{gX}(t), D_{gX}(s)\} = \{2\pi f_X(0)\}^2 \mathit{\Gamma}_g(s,t)$$

where $\mathit{\Gamma}_g(s,t)$ is as defined by (14).

We denote the corresponding zero mean process by $\tilde{D}_{gX}(t)$; that is,

$$\tilde{D}_{gX}(t) = D_{gX}(t) - \{2\pi f_X(0)\} \int_0^t g(x,x)\,dx \quad .$$

The covariance kernel for $\{\tilde{D}_g(t),\ t \in [0,1]\}$ of course remains as that given in Theorem 3. Effects due to non-normality of the error variables disappear for large samples.

5 The Squared Residuals Process

We return to consideration of the partial sums of squared residuals as defined in Section 2. Equation (5) defines these partial sums in terms of the processes discussed in the two previous sections. Since

$$\chi_n^2(t) - \rho_{gn}(t) = n^{-1/2} D_{gn}(t)$$

and since $n^{-1/2} \sup\limits_{0\le t\le 1} |D_{gn}(t)| \longrightarrow 0$ with probability 1, the large sample distribution theory for $\{\tilde{\rho}_{gn}(t),\ t \in [0,1]\}$ is the same as that for $\{\tilde{\chi}_n(t),\ t \in [0,1]\}$ where $\tilde{\rho}_{gn}(t) = \rho_{gn}(t) - E\{\rho_{gn}(t)\}$.

To account for the effect of regression, the finite sample correction factor for the mean is approximated by

$$n^{-1/2}\{2\pi f_X(0)\}\int_0^t g(x,x)\,dx \ \ ;$$

that is,

$$E\{\rho_{gX_n}(t)\} = \sqrt{n}\,tR_X(0) - \frac{2\pi f_X(0)}{\sqrt{n}}\int_0^t g(x,x)\,dx + O(1/n). \tag{15}$$

The finite sample covariance kernel is approximated as follows:

$$\begin{aligned} Cov\{\rho_{gX_n}(t), \rho_{gX_n}(s)\} &= K_X(s\wedge t) - \frac{1}{n}\{2\pi f_X(0)\}^2 \Gamma_g(s,t) \\ &\quad -\frac{F_{X2}}{n}\Gamma_g(s,t) + O(1/n^2) \end{aligned} \tag{16}$$

where

$$F_{X2} = \sum_{|\nu_i|<\infty}\sum C_{X4}(0,\nu_1,\nu_2). \tag{17}$$

It can be shown that

$$Cov\left\{\sqrt{n}\rho_{gX_n}(t), D_{gX_n}(s)\right\} = \frac{F_{X_2}}{2}\Gamma_g(s,t) + O(1/n) \ \ .$$

Thus the two processes $\{\rho_{gX_n}(t),\ t\in[0,1]\}$ and $\{D_{gX_n}(t),\ t\in[0,1]\}$ are approximately uncorrelated if the higher order cumulants are zero. Of course this holds exactly if in addition there is no serial correlation in the noise process and if $t=1$.

6 Cumulants for Linear Time Series

We now discuss the nature of cumulant expressions such as those appearing in formulae (11) and (17) for the case of linear time series defined as follows:

$$X(t) = \sum_{j=0}^{\infty} a_j \epsilon(t-j) \qquad t = 0, \pm1, \ldots$$

where $a_0 = 1$, $\sum_{j=0}^{\infty}|a_j| < \infty$ and the innovations $\epsilon(j)$, $(j = 0, \pm1, \ldots)$ form a sequence of independent and identically distributed random variables such that $E\{\epsilon(j)\} = 0$, $E\{\epsilon^2(j)\} = \sigma^2 < \infty$ and $\mathrm{Cum}\{\epsilon^4(j)\} = k_4$, $|k_4| < \infty$. It may be shown that:

$$F_{X1} \ = \sum_{|\nu|<\infty} C_{X4}(0,\nu,\nu) = k_4\left(\sum_{j=0}^{\infty} a_j^2\right)^2 = \frac{k_4}{\sigma^4}R_X^2(0) \ \ ,$$

$$F_{X2} = \sum_{|\nu_i|<\infty}\sum C_{X4}(0,\nu_1,\nu_2) = k_4\left(\sum_{j=0}^{\infty} a_j^2\right)\left(\sum_{j=0}^{\infty} a_j\right)^2 = \frac{k_4}{\sigma^4}R_X(0)\{2\pi f_X(0)\}$$

and

$$F_{X3} = \sum_{|\nu_i|<\infty}\sum\sum C_{X4}(\nu_1,\nu_2,\nu_3) = k_4\left(\sum_{j=0}^{\infty} a_j\right)^4 = \frac{k_4}{\sigma^4}\{2\pi f_X(0)\}^2 \quad .$$

In the event of white noise error structure then $F_{X1} = F_{X2} = F_{X3} = k_4$.

7 Variable Sampling Rates

As indicated in Section 2, the time of observation need not be restricted to $[0,1]$ nor must sampling be equi-spaced for the above results to apply. Suppose the total sampling period is $[0,T]$ and the rate of sampling, or the density of observations, is described by a non-constant, positive function $\{s(t),\ t \in [0,T]\}$ Riemann integrable to 1 on the interval $[0,T]$. If $S(t) = \int_0^t s(x)\,dx$, then $S(t_2) - S(t_1)$ is the proportion of the observations in the interval $[t_1,t_2]$. Also, if

$$\sqrt{n}\chi^2_{XS_n}(t) = \sum_{j=1}^{[nS(t)]} X^2(j) + \{nS(t) - [nS(t)]\}^2\, X^2([nS(t)]+1) \quad ,$$

then the mean corrected process $\tilde{\chi}^2_{XS_n}(\cdot)$ converges weakly to the zero-mean, gaussian process $B_{X^2S}(\cdot)$ which has covariance kernel $K_{X^2S}(s,t) = K_X\{S(s)\wedge S(t)\}$. But the covariance kernel of the limit process $B_{X^2}(\cdot)$ is such that $K\{S(s),S(t)\} = K_X\{S(s)\wedge S(t)\}$. Hence, $B_{X^2S}(t)$ and $B_{X^2}\{S(t)\}$ are stochastically equivalent.

Also, if the vector of regressor functions, $\mathbf{g}(\cdot)$, is defined on the interval $[0,T]$, then expression (14) becomes

$$\varGamma_{gS}(s,t) = 4\int_0^{s\wedge t} g_S(x,x)\,dS(x) - 2\int_0^t\int_0^s g_S^2(x,y)\,dS(x)\,dS(y).$$

where $g_S(x,y) = \boldsymbol{g}'(x)\boldsymbol{G}_S^{-1}\boldsymbol{g}(y)$ and the i,jth component of $\boldsymbol{G}_S$ is

$$\boldsymbol{G}_{Sij} = \int_0^T g_i(x)g_j(x)\,dS(x) \quad .$$

The transformations $u = S(x)$ and $v = S(y)$ can be used to show that, for $0 \le s,t \le 1$, $\varGamma_{gS}\{S^{-1}(s),S^{-1}(t)\} = \varGamma_{g^*}(s,t)$ where the bilinear form in (14) becomes $\boldsymbol{g}^*(u,v) = \boldsymbol{g}^{*\prime}(u)\boldsymbol{G}^{*-1}\boldsymbol{g}^*(v)$ with $\boldsymbol{g}^*(u) = \boldsymbol{g}\{S^{-1}(u)\}$ and $\boldsymbol{G}^*_{ij} = \int_0^1 g_i^*(u)g_j^*(u)\,du$.

Thus the covariance functions for the limit processes for the variable sampling rates case are related to those discussed above for the the case of equispaced sampling on $[0,1]$ as follows:

i) $Cov[\chi^2_{XS}\{S^{-1}(t)\}, \chi^2_{XS}\{S^{-1}(s)\}] = Cov\{\chi^2_X(t), \chi^2_X(s)\}$,

ii) $Cov[D_{gXS}\{S^{-1}(t)\}, D_{gXS}\{S^{-1}(s)\}] = Cov\{D_{g^*X}(t), D_{g^*X}(s)\}$
and

iii) $Cov[\rho_{gXS}\{S^{-1}(t)\}, \rho_{gXS}\{S^{-1}(s)\}] = Cov\{\rho_{g^*X}(t), \rho_{g^*X}(s)\}$.

Note that the noise process is assumed to form a stationary series regardless of the sampling rate. Care must be taken in the estimation of $f(0)$ to avoid the biases produced by the regression fit extracting low frequency power from the spectrum of the noise process (see Tang and MacNeill (1993)) and by unequal sampling intervals (see Jones (1985)).

8 Examples

A. Polynomial Regression

We consider polynomial regression where $g_i(t) = t^i$ $(i = 0, 1, \ldots, p)$. Jandhyala and Minogue (1993) have shown that

$$g(s,t) = \sum_{m=0}^{p} (2m+1)\gamma_m(s)\gamma_m(t)$$

where

$$\gamma_m(t) = \sum_{j=0}^{m} (-1)^{m+j} \begin{pmatrix} m+j \\ j\ j\ m-j \end{pmatrix} t^j \quad .$$

Hence for $p = 0$, $g(x,y) \equiv 1$. For $p = 1$, $g(x,y) = 1 + 3(1-2x)(1-2y)$, for $p = 2$, $g(x,y) = 1 + 3(1-2x)(1-2y) + 5(1-6x+6x^2)(1-6y+6y^2)$, etc.

The crucial function defined by the regression model for evaluating first and second moments of squared residuals processes is $\Gamma_g(s,t)$ as defined by (14). This is a bilinear form in powers of s and t, and hence can be expressed as follows: $\Gamma_g(s,t) = \boldsymbol{s}'\boldsymbol{\Gamma}\boldsymbol{t}$. For the case $p = 0$, if $\boldsymbol{s}' = (1, s)$, $\boldsymbol{t}' = (1, t)$ and $t < s$, then

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & 4 \\ 0 & -2 \end{pmatrix}$$

The transpose of this matrix should be used if $s < t$. Then $\Gamma_g(s,t) = 4t - 2st$ and $\Gamma_g(1,1) = 2$.

For the case $p = 1$, if $\boldsymbol{s}' = (1, s, s^2, s^3)$, $\boldsymbol{t}' = (1, t, t^2, t^3)$ and $t < s$, then $\Gamma_g(s,t) = \boldsymbol{s}'\boldsymbol{\Gamma}\boldsymbol{t}$ where

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & 16 & -24 & 16 \\ 0 & -32 & 48 & -24 \\ 0 & 48 & -84 & 48 \\ 0 & -24 & 48 & -32 \end{pmatrix}$$

It can be shown that $\Gamma_g(1,1) = 4$.

For the case $p = 2$, with $\boldsymbol{s}' = (1, s, \ldots, s^5)$, $\boldsymbol{t}' = (1, \ldots, t^5)$ and $t < s$ then $\Gamma_g(s,t) = \boldsymbol{s}'\boldsymbol{\Gamma}\boldsymbol{t}$ where

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & 36 & -144 & 336 & -360 & 144 \\ 0 & -162 & 648 & -1224 & 1080 & -360 \\ 0 & 648 & -3024 & 6408 & -6120 & 2160 \\ 0 & -1224 & 6408 & -15072 & 15480 & -5760 \\ 0 & 1080 & -6120 & 15480 & -16740 & 6480 \\ 0 & -360 & 2160 & -5760 & 6480 & -2592 \end{pmatrix}$$

We note that $\Gamma_g(1,1) = 6$. Computer algebra systems can be used to evaluate $\boldsymbol{\Gamma}$ for higher order polynomials.

As a specific example we consider simple linear regression with an MA(2) noise process defined in terms of uniform errors on the interval $[-\theta/2, \theta/2]$. Then: $\sigma^2 = \theta^2/12$, $k_4 = -\theta^4/120$, $F_{X1} = k_4(1+a_1^2+a_2^2)^2$, and $F_{X2} = k_4(1+a_1^2+a_2^2)(1+a_1+a_2)^2$. For $t < s$,

$$\begin{aligned} \Gamma_g(s,t) &= 16t - 24t^2 + 16t^3 + 32st + 24s^3t + 245st^3 - 48s^2t - 48st^2 \\ &\quad -48s^3t - 48s^2t^3 + 845s^2t^2 + 325s^3t^3 \quad . \end{aligned}$$

Also

$$4\pi \int_{-\pi}^{\pi} f_X^2(\lambda)\, d\lambda = 2\sigma^4 \left(1 + a_1^4 + a_2^4 + 4a_1^2 + 4a_2^2 + 4a_1^2 a_2 + 4a_1^2 a_2^2\right) \quad ,$$

and

$$2\pi f_X(0) = \sigma^2(1 + a_1 + a_2)^2 \quad .$$

Then from (15) and (16) one can compute the means and covariances of the partial sums of regression residuals.

If sampling is sparse at first and more dense later, and is characterized by $S(t) = t^2, 0 \le t \le 1$, then the covariances of the partial sums can be computed using $g(x,y) = 9 - 12x - 12y + 18xy$. Then, for $t < s$,

$$\begin{aligned} \Gamma_{gS}(s,t) &= 36t^2 - 64t^3 + 36t^4 - 18s^2t^2 + 32s^2t^3 - 16s^2t^4 + 32s^3t^2 - (544/9)s^3t^3 \\ &\quad +32s^3t^4 - 16s^4t^2 + 32s^4t^3 - 18s^4t^4 \ , \end{aligned}$$

and $\Gamma_{gS}(\sqrt{s}, \sqrt{t}) = \Gamma_{g^*}(s,t)$.

B. Harmonic Regression

As another example we consider harmonic regression with $g_k(t) = \cos 2\pi kt$ $(k = 0, 1, \ldots, p)$ and $g_k(t) = \sin 2\pi kt$ $(k = p+1, \ldots, 2p)$. It was shown by MacNeill (1978b) that

$$g(s,t) = 1 + 2\sum_{j=1}^{p} \{\sin 2\pi jt \sin 2\pi js + \cos 2\pi js \cos 2\pi jt\} \quad .$$

If the error process is normal white noise then it may be shown that

$$\sigma^{-2} E\{D_g(t)\} = (2p+1)t$$

and

$$\sigma^{-4} Cov\{D_g(t), D_g(s)\} = 4(2p+1)\min(s,t) - 2\int_0^t\int_0^s \left\{1 + 2\sum_{j=1}^{p} \cos 2\pi j(x-y)\right\}^2 dxdy$$

We note that $E\{D_g(1)\} = (2p+1)\sigma^2$, $\mathrm{Var}[D_g(1)] = 2(2p+1)\sigma^4$ and $\sigma^{-2}D_g(1)$ is χ^2 distributed on $2p+1$ degrees of freedom. Furthermore,

$$E\left\{\sum_{j=1}^{[nt]} r_{gX_n}^2(j)\right\} \simeq \sigma^2 t\{n - (2p+1)\}$$

and

$$\begin{aligned} \mathrm{Var}\left\{\sum_{j=1}^{[nt]} r_{gX_n}^2(j)\right\} &\simeq \mathrm{Var}\left\{\sum_{j=1}^{[nt]} \epsilon^2(j)\right\} - \mathrm{Var}\{D_g(t)\} \\ &\simeq 2\sigma^4 nt - 4(2p+1)\sigma^4 t \\ &\quad + 2\sigma^4 \int_0^t \int_0^s \left\{1 + 2\sum_{j=1}^{p} \cos 2\pi j(x-y)\right\}^2 dx\,dy \ . \end{aligned}$$

For the particular model with $p = 1$, $g(x,z) = 1 + 2\cos 2\pi(x-z)$, and

$$\begin{aligned} \mathit{\Gamma}_g(s,t) &= -\frac{2}{\pi^2}\cos\{2\pi(s-t)\} + \frac{\sin^2\{2\pi(s-t)\}}{2\pi^2} + \frac{2\cos 2\pi s}{\pi^2} \\ &\quad -\frac{\sin^2(2\pi s)}{2\pi^2} + \frac{2\cos(2\pi t)}{\pi^2} - \frac{\sin^2(2\pi t)}{2\pi^2} - \frac{2}{\pi^2} - 6st + 12t \ , \end{aligned}$$

with $\mathit{\Gamma}_g(1,1) = 6$.

9 A Change-Point Statistic Based on Cusums of Squares of Raw Residuals

A statistic proposed by Hsu (1977) for testing for constancy of variance involves the sequence of partial sums of squares of residuals normalized by what amounts to an estimate of the variance. In our context this amounts to consideration of $\rho_{gX_n}(t)/\rho_{gX_n}(1)$.

It can be shown that the ratio

$$\sqrt{n}\frac{\rho_{gX_n}(t)}{\rho_{gX_n}(1)} \doteq \frac{1}{\hat{\sigma}_n^2\sqrt{n}}\sum_{j=1}^{[nt]} r_{gX_n}^2(j)$$

when properly normalized converges weakly to a Brownian Bridge. That is,

$$\frac{1}{\sqrt{n}\hat{\sigma}_n^2}\sum_{j=1}^{[nt]} r_{gX_n}^2(j) - \sqrt{n}t \Longrightarrow \left(\frac{K_X}{R_X^2(0)}\right)^{1/2} B_0(t). \tag{18}$$

Since $K_X = 4\pi\int_{-\pi}^{\pi} f_X^2(\lambda)\,d\lambda + F_{X1}$, one can determine the effects of serial correlation and non-normality on the large sample distribution of the statistic. As discussed above, these effects can be substantial.

Change-point statistics can be obtained by defining appropriate functionals on (18). Large sample distributional results for a variety of Cramér-von Mises type statistics are presented by Tang and MacNeill (1992).

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The author wishes to thank the editor for the opportunity to express his appreciation to Professor Safiul Haq for the many happy years of collegiality they enjoyed.

References

[1] Billingsley, P. (1968). *Convergence of Probability measures.* Wiley, New York.

[2] Brillinger, D.R. (1973). "Estimation of the mean of a stationary time series by sampling". *Journal of Applied Probability* **10,** pp. 419–431

[3] Chernoff, H. and S. Zacks (1964). "Estimating the current mean of a distribution which is subject to changes in time". *Annals of Mathematical Statistics* **35,** pp. 999–1018.

[4] De Gooijer, J.G. and I.B. MacNeill (1999). "Lagged regression residuals and serial correlation tests". *Journal of Business and Economic Statistics* **17,** pp. 236–247.

[5] Harrison, M.J. and B.P.M. McCabe (1979). "A test for heteroscadisticity based on ordinary least squares residuals". *Journal of the American Statistical Association* **74** pp. 494–499.

[6] Hsu, D.A. (1977). "Tests for variance shift at an unknown time point". *Applied Statistics* **26,** pp. 279–284.

[7] Inclán, C. and G.C. Tiao (1994). "Use of cumulative sums of squares for retrospective detection of changes of variance". *Journal of the American Statistical Association* **89,** pp.913–923.

[8] Jandhyala, V.K. and I.B. MacNeill (1989). "Residual partial sum limit processes for regression models with application to detecting parameter changes at unknown times". *Stochastic Processes and their Applications* **33,** pp. 309–323.

[9] Jandhyala, V.K. and C.D. Minogue (1993). "Distributions of Bayes-type change-point statistics under polynomial regression". *Journal of Statistical Planning and Inference* **37** pp.271–290.

[10] Jones, R.H. (1985). "Fitting multivariate models to unequally spaced data. In *Time Series Analysis of Irregularly Observed Data*, ed. E. Parzen. Berlin: Springer–Verlag, pp. 158–188.

[11] MacNeill, I.B. (1978a). "Properties of partial sums of polynomial regression residuals with application to tests for change in regression at unknown times". *Annals of Statistics* **6,** pp. 422–433.

[12] MacNeill, I.B. (1978b). "Limit processes for sequences of partial sums of regression residuals". *Annals of Probability* **6,** pp. 695–698.

[13] Tang, S.M. and I.B. MacNeill (1992). "Monitoring statistics which have increased power over a reduced time range". *Environmental Monitoring and Assessment* **23,** pp. 189–203.

[14] Tang, S.M. and I.B. MacNeill (1993). "The effect of serial correlation on tests for parameter change at unknown time". *Annals of Statistics* **21,** pp. 552–575.

[15] Tsay, R. (1988). "Outliers, level shifts, and variance changes in time series". *Journal of Forecasting* **7,** pp. 1–20.