The Annals of Statistics
1993, Vol. 21, No. 1, 552–575

THE EFFECT OF SERIAL CORRELATION ON TESTS FOR PARAMETER CHANGE AT UNKNOWN TIME[1]

BY S. M. TANG AND I. B. MACNEILL

Compusearch Market and Social Research Ltd. and University of Western Ontario

It is shown that serial correlation can produce striking effects in distributions of change-point statistics. Failure to account for these effects is shown to invalidate change-point tests, either through increases in the type 1 error rates if low frequency spectral mass predominates in the spectrum of the noise process, or through diminution of the power of the tests when high frequency mass predominates. These effects are characterized by the expression $\{2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda\}$, where $f(\cdot)$ is the spectral density of the noise process; in sample survey work this is known as the design effect or "deff." Simple precise adjustments to change-point test statistics which account for serial correlation are provided. The same adjustment applies to all commonly used regression models. Residual processes are derived for both stationary time series satisfying a moment condition and for general linear regression models with stationary error structure.

1. Introduction. Stochastic models for time sequenced data are generally characterized by several unknown parameters. These parameters may change over time, and if the changes, when they occur, do so unannounced and at unknown time points, then the associated inferential problem is referred to as the change-point problem. Various important application areas of statistics involve change detection in a central way; two of these areas are quality assurance and environmental monitoring.

Most of the statistics commonly applied to the change-point problem involve cumulative sums or partial sums of regression residuals. The distribution theory for these statistics has been computed under the assumption that the error process for the regression model is white noise. In this paper we consider linear regression of a random variable against general nonstochastic functions of time, but with error variables that form a serially correlated time series. We then examine the large sample properties of the stochastic processes defined by the partial sums of the regression residuals. Large sample distribution theory for fixed sample size statistics used to detect changes in regression parameters is usually derived by computing the distributions of various functionals on

Received October 1989; revised March 1992.

[1]Partially supported by a contract from the Ontario Ministry of Industry Trade and Technology, and a grant from the Natural Sciences and Engineering Research Council of Canada.

AMS 1991 *subject classifications*. Primary 62M10; secondary 62E20, 62G10, 62J05, 62M15.

Key words and phrases. Serial correlation, cumulative sums, change-point statistics, spectral density, residuals processes, partial sums.

these limit processes. Our results given below show that serial correlation in the error process can produce striking effects in these distributions.

Among the statistics of concern to us are those derived through a Bayesian-type approach introduced to the change-point problem by Chernoff and Zacks (1964). Application of this method by Jandhyala and MacNeill (1989, 1991) yielded Cramér–von Mises type statistics defined in terms of the sequence of partial sums of regression residuals. The residual process for general linear regression with white noise errors was derived by MacNeill (1978a, b). MacNeill and Jandhyala (1985) examined the properties of the residual process for nonlinear regression. Johnson and Bagshaw (1974) obtained the limit processes for partial sums of observations from ARMA processes and explored the effect of ARMA noise on the CUSUM statistics proposed by Page (1954). In a sequel, Bagshaw and Johnson (1975) examined the effect of ARMA noise on the run length distribution for CUSUM's. Kulperger (1985) considered polynomial regression and extended the results of MacNeill (1978a) to autoregressive error processes.

Likelihood ratio statistics for testing for changes in mean and in simple linear regression in the presence of white noise have been discussed by Quandt (1960), Hawkins (1977), Worsley (1979, 1983), James, James and Siegmund (1987) and Kim and Siegmund (1989). These statistics are based on cumulative sums of residuals scaled by their standard deviations before maximization. In addition to providing test statistics, this approach permits joint inferences regarding the change-point and the parameters. Worsley (1986) and Siegmund (1988) discuss confidence sets for the change-point problem.

Brown, Durbin and Evans (1975) proposed ad hoc procedures for testing for parameter change at unknown time in regression parameters. Their procedures are based on recursive residuals, and are capable of being applied to regression on a wide class of nonstochastic regressors.

Picard (1985) uses likelihood based methods to derive statistics for testing for parameter change at unknown time in autoregressive models. These methods are extended to the problem of detecting possible changes occurring at the same instant in both a mean and autoregressive parameters.

El-Shaarawi and Esterby (1982) considered inferences about the point of change in regression models with AR(1) error structure. Brillinger (1989) considered the problem of consistent detection of a monotonic trend superimposed on a stationary time series.

The plan of this paper is as follows. After introducing the problem in Section 2 we derive in Section 3 the residual processes for stationary time series satisfying a moment condition. These processes are used in Section 4 to obtain the residual processes for regression against general nonstochastic regression functions of time when the errors form a stationary time series. After a discussion of several examples, we indicate in Section 6 how the properties of the residual process can be used to adjust distributional results for a wide class of statistics defined in terms of partial sums of residuals to account for serial correlation. We discuss and illustrate adjustment factors for

common noise models. We then discuss in Section 8 the effect of serial correlation on change-point statistics and provide simple, easily applied, large sample adjustments to account for this serial correlation. Finally, in Section 9 we present the results of a simulation which give some indication of the circumstances to which the asymptotic results apply and of those to which they should be applied with some caution.

2. Regression models and error process structure. We first define the basic model. Let $X(n)$, $n = 0, \pm 1, \ldots$, be a zero mean, discrete time, stationary time series with covariance function $R(\nu) = E\{X(n)X(n+\nu)\}$, $|\nu| < \infty$. If the covariance function is absolutely summable, that is,

$$\sum_{\nu=-\infty}^{\infty} |R(\nu)| < \infty, \tag{1}$$

then the spectral density function, $f(\lambda) = (1/2\pi)\sum_{|\nu|<\infty} e^{-i\lambda\nu}R(\nu)$, $\lambda \in [-\pi, \pi]$, exists. If the spectral density is positive, that is, if

$$f(\lambda) \geq a > 0, \qquad \lambda \in [-\pi, \pi], \tag{2}$$

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. Several different sets of conditions guarantee convergence in distribution of $n^{-1/2}\sum_{j=1}^{[nt]} X(j)$, $t \in [0, 1]$, ($[nt]$ is the integer part of nt) to the normal with zero mean and variance $\{2\pi f(0)t\}$; these include those given by Hannan (1973) and those given by Brillinger (1973). Neither set of conditions implies the other, although both include most processes encountered in practice. Since the Brillinger conditions are more easily stated we use them. They are stated in terms of cumulant functions, which are defined as follows:

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

Stationarity to order $k+1$ is implicit in this definition. Second order stationarity together with finiteness of the variance implies weak stationarity. Note that the first two cumulants are $E\{X(n)\}$ and $R(\nu)$, $|\nu| < \infty$. When necessary we assume the cumulants exist and satisfy what we call the Brillinger conditions, namely,

$$|C_{k+1}(\nu_1, \nu_2, \ldots, \nu_k)| < \frac{L_k}{\prod_{j=1}^{k}(1+\nu_j^2)} \tag{3}$$

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

We now consider the regression part of the model and let $\{g_k(\cdot), 0 \leq k \leq p\}$ be a collection of regressor functions defined on the unit interval $[0, 1]$. A triangular array $Y_n(j)$, $1 \leq j \leq n$, $n \geq 1$, of dependent variables is defined as

follows:

$$Y_n(j) = \sum_{i=0}^{p} \beta_i g_i\left(\frac{j}{n}\right) + X(j). \tag{4}$$

For convenience only, 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. The matrix formulation of this model is

$$\mathbf{Y}_n = \mathbf{A}_n \beta_p + \mathbf{X}_n, \tag{5}$$

where $\mathbf{X}'_n = \{X(1), X(2), \ldots, X(n)\}$ is a portion of a realization of the stationary time series and where the (j, i)th component of the design matrix is $g_i(j/n)$. The regression parameter estimators are denoted by $\tilde{\beta}_{pn}$ where

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

Least squares estimators were shown by Grenander (1954) to be efficient under quite general error process assumptions, although for purposes of inference it must be kept in mind that the covariance matrix for the parameters depends upon the spectrum of the noise process.

Sequences of partial sums of regression residuals are denoted by $Sg_{nj} (1 \leq j \leq n,\ n \geq 1)$, where $Sg_{nj} = \Sigma_{i=1}^{j} \{Y_n(i) - \hat{Y}_n(i)\}$, $\hat{Y}_n(i) = \tilde{\beta}'_{pn} g(i/n)$ and $g'(i/n) = \{g_0(i/n), g_1(i/n), \ldots, g_p(i/n)\}$. $Sg_{n0} \equiv 0$. 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 to define a sequence of stochastic processes $[\{\theta_{gX_n}(t),\ t \in [0, 1]\},\ n \geq 1]$ possessing continuous sample paths as follows:

$$n^{1/2}\theta_{gX_n}(t) = Sg_{n[nt]} + (nt - [nt])\{Y([nt] + 1) - \hat{Y}([nt] + 1)\}.$$

Then if $\mathbf{e}_{n,nt}$ is an $n \times 1$ vector whose first $[nt]$ components are 1, the next is $nt - [nt]$ and the remainder zero, one can write

$$n^{1/2}\theta_{gX_n}(t) = \mathbf{e}'_{n,nt}\{\mathbf{I}_n - \mathbf{A}_n(\mathbf{A}'_n \mathbf{A}_n)^{-1} \mathbf{A}'_n\}\mathbf{X}_n. \tag{6}$$

Provided the Riemann integrals on [0, 1] of $g_r^2(\cdot)$, $r = 0, \ldots, p$, exist then the (r_1, r_2)th component of $\lim n^{-1}(\mathbf{A}'_n \mathbf{A}_n) \equiv \mathbf{G}$ is $\int_0^1 g_{r_1}(t) g_{r_2}(t)\, dt$. We then define $g(s, t) \equiv \mathbf{g}'(s)\mathbf{G}^{-1}\mathbf{g}(t)$, assuming, here and in the sequel, that the inverse in this bilinear form exists.

3. The partial sum limit process for stationary time series. To establish the limit process for $[\{\theta_{gX_n}(t),\ t \in [0, 1]\},\ n \geq 1]$ we need first to examine the properties of the sequence of partial sums of the error process $X(n)$, $n = 0, \pm 1, \ldots$. Hence we let $S_{X_j} = \Sigma_{i=1}^{j} X(i)$ and define another sequence of stochastic processes $[\{\theta_{X_n}(t),\ t \in [0, 1]\},\ n \geq 1]$ possessing continuous sample paths by

$$n^{1/2}\theta_{X_n}(t) = S_{X_{[nt]}} + (nt - [nt])X([nt] + 1). \tag{7}$$

We note first that $\theta_{X_n}(0) = E\{\theta_{X_n}(t)\} \equiv 0$ and consider next the covariance

kernel of the process, $K(s,t) = E\{\theta_{X_n}(s)\theta_{X_n}(t)\}$. We assume $t = \min(s,t)$, $k = [nt]$ and $l = [ns]$. Since

$$\left| E\{\theta_{X_n}(s)\theta_{X_n}(t)\} - E\left\{\theta_{X_n}\left(\frac{k}{n}\right)\theta_{X_n}\left(\frac{l}{n}\right)\right\}\right| \leq \frac{c}{n}, \tag{8}$$

where $c > 0$ is independent of s, t and n, for large samples we need only consider $K_n(k/n, l/n)$. Then

$$\begin{aligned} K_n\left(\frac{k}{n}, \frac{l}{n}\right) &= \frac{1}{n}E(S_{X_k}S_{X_l}) \\ &= \frac{1}{n}\sum_{t_1=1}^{k}\sum_{t_2=1}^{l} E\{X(t_1)X(t_2)\} \\ &= \frac{1}{n}\sum_{t_1=1}^{k}\sum_{t_2=1}^{l} R(t_2 - t_1). \end{aligned}$$

Hence the covariance kernel of the process is defined in terms of the sum of the elements of the covariance function $R(t_2 - t_1)$ for (t_1, t_2) in the region of Figure 1 defined by $A \cup B \cup C$; that is

$$K_n\left(\frac{k}{n}, \frac{l}{n}\right) = \frac{1}{n}\left\{\sum_{(t_1,t_2)\in B\cup D} + \sum_{(t_1,t_2)\in A\cup C} - \sum_{(t_1,t_2)\in D}\right\}R(t_2 - t_1).$$

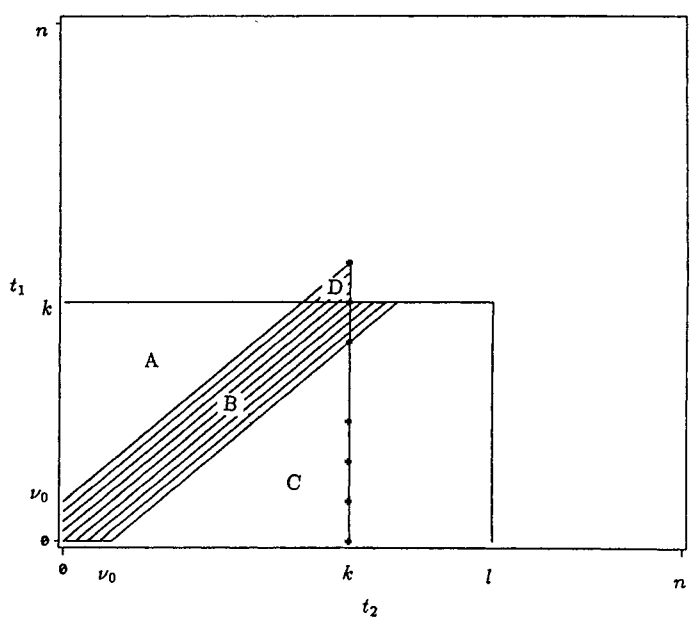


FIG. 1. *Schematic diagram of the elements of the covariance kernel* $K(s,t)$.

Since

$$\sum_{(t_1,t_2)\in B\cup D} R(t_1-t_2) = k\sum_{|\nu|\le\nu_0} R(\nu),$$

$$\left|\sum_{(t_1,t_2)\in D} R(t_1-t_2)\right| = \left|\sum_{\nu=1}^{\nu_0} \nu R(-\nu)\right| \le \nu_0 \sum_{\nu=1}^{\nu_0} |R(-\nu)|$$

and

$$\left|\sum_{(t_1,t_2)\in A\cup C} R(t_2-t_1)\right| \le k\sum_{\nu_0<|\nu|\le l} |R(\nu)| \le k\sum_{|\nu|>\nu_0} |R(\nu)|,$$

then if the covariance function satisfies (1), for any $\varepsilon > 0$ we can choose ν_0 such that $\sum_{|\nu|>\nu_0}|R(\nu)| < \varepsilon$, and also, having selected ν_0, we can choose n_0 such that if $n > n_0$, then

$$\frac{\nu_0}{n}\sum_{\nu=1}^{\nu_0} |R(-\nu)| < \varepsilon.$$

Hence,

$$\begin{aligned}\left|K_n\left(\frac{k}{n},\frac{l}{n}\right) - \frac{k}{n}\sum_{|\nu|<\infty} R(\nu)\right| &\le \left|\frac{k}{n}\sum_{|\nu|>\nu_0} R(\nu)\right| + \left|\frac{1}{n}\sum_{(t_1,t_2)\in D} R(t_2-t_1)\right| \\ &\quad + \left|\frac{1}{n}\sum_{(t_1,t_2)\in A\cup C} R(t_2-t_1)\right| \\ &\le \frac{k}{n}\varepsilon + \frac{\nu_0}{n}\sum_{\nu=1}^{\nu_0}|R(-\nu)| + \frac{k}{n}\sum_{|\nu|>\nu_0}|R(\nu)| \\ &\le 3\varepsilon.\end{aligned}$$

If we combine this with (8) and recall that $2\pi f(0) = \sum_{|\nu|<\infty} R(\nu)$, we see that

$$K_n(s,t) \to 2\pi f(0)\min(s,t)$$

uniformly in (s,t).

If $t_1 < t_2 < t_3$, an argument similar to that used above can be used to show

$$E\left[\left\{\theta_{X_n}(t_2) - \theta_{X_n}(t_1)\right\}\left\{\theta_{X_n}(t_3) - \theta_{X_n}(t_2)\right\}\right] \to 0 \quad \text{as } n \to \infty.$$

This implies the covariance matrix for the p-vector $\{\theta_{X_n}(t_1),\ldots,\theta_{X_n}(t_p)\}$ is $\|2\pi f(0)\min(t_i,t_j)\|$. We establish asymptotic normality in the following theorem.

THEOREM 1. *Under assumptions* (2) *and* (3) *the p-vector* $\{\theta_{X_n}(t_1),\ldots,\theta_{X_n}(t_p)\}$ *has a nontrivial asymptotic probability distribution that is normal with zero mean and covariance matrix* $\|2\pi f(0)\min(t_i,t_j)\|$.

PROOF. The Brillinger (1973) condition (3) guarantees univariate asymptotic normality of the components of the vector, and the Cramér–Wold device

of demonstrating asymptotic multivariate normality by showing asymptotic normality with zero mean and variance $2\pi f(0)\sum_{j=1}^{p}(\sum_{i=j}^{p}\lambda_i)^2(t_j - t_{j-1})$ of $\sum_{j=1}^{p}\lambda_j\theta_{X_n}(t_j)$, $0 = t_1 < t_2 < \cdots < t_n$, where the λ_j are arbitrary real numbers, can be used to complete the proof for the p-dimensional case. □

We next show tightness of the sequence of measures P_{X_n}, $n = 1, 2, \ldots$, generated in $C[0, 1]$ by $[\{\theta_{X_n}(t),\ t \in [0, 1]\},\ n \geq 1]$. The arguments used above to derive the covariance kernel for these processes can be used to show the existence of a constant $C > 0$ such that for $t_2 > t_1$,

$$E\{\theta_{X_n}(t_2) - \theta_{X_n}(t_1)\}^4 \leq C(t_2 - t_1)^2, \tag{9}$$

where C is not dependent on t_1, t_2 nor n. To demonstrate (9) we assume n large and approximate $\theta_{X_n}(t_2) - \theta_{X_n}(t_1)$ by $(1/\sqrt{n})S_l - (1/\sqrt{n})S_k$ where $[nt_1] = k$ and $[nt_2] = l$. Then

$$E\left\{\frac{1}{\sqrt{n}}S_l - \frac{1}{\sqrt{n}}S_k\right\}^4 = \frac{1}{n^2}\sum_{(j_1, j_2, j_3, j_4)\in A} E\{X(j_1)X(j_2)X(j_3)X(j_4)\},$$

where $A = \{(j_1, j_2, j_3, j_4)\colon k < j_i \leq l,\ i = 1, \ldots, 4\}$. We now partition A as in Figure 1, except we allow for four indices instead of two. We let

$$B = \{(j_1, j_2, j_3, j_4)\colon j_2 = j_1 + \nu_1,\ j_3 = j_1 + \nu_2,\ j_4 = j_1 + \nu_3,\ \text{at least one } j_i \in (k, l],\ |\nu_i| \leq \nu_0\}$$

and

$$C = \{(j_1, j_2, j_3, j_4)\colon j_2 = j_1 + \nu_1,\ j_3 = j_1 + \nu_2,\ j_4 = j_1 + \nu_3,\ k < j_i \leq l,\ i = 1, 2, 3, 4,\ |\nu_i| > \nu_0\}.$$

We also define as follows a set D which lies outside A:

$$D = \{(j_1, j_2, j_3, j_4)\colon j_2 = j_1 + \nu_1,\ j_3 = j_1 + \nu_2,\ j_4 = j_1 + \nu_3,\ \text{at least one } j_i \in (k, l],\ \text{at least one } j_i \notin (k, l]\}.$$

These fourth order moments can be expressed in terms of the corresponding fourth order cumulants and products of pairs of elements from the covariance function. Hence,

$$\begin{aligned}\frac{1}{n^2}\sum_A E\left\{\prod_{i=1}^{4}X(j_i)\right\} &= \frac{1}{n^2}\sum_A \operatorname{Cum}\{X(j_1), X(j_2), X(j_3), X(j_4)\} \\ &\quad + \frac{1}{n^2}\sum_A \{R(j_1 - j_2)R(j_3 - j_4) + R(j_1 - j_3)R(j_2 - j_4) \\ &\qquad + R(j_1 - j_4)R(j_2 - j_3)\}.\end{aligned}$$

However,

$$\frac{1}{n^2}\sum_A R(j_1-j_2)R(j_3-j_4) = \left\{\frac{1}{n}\sum_{l<j_1\le k}\sum_{l<j_2\le k} R(j_1-j_2)\right\}^2,$$

which was shown above to converge to $\{2\pi f(0)(t_2-t_1)\}^2$. We now consider the fourth order cumulant terms:

$$\frac{1}{n^2}\sum_A \operatorname{Cum}\{X(j_1), X(j_2), X(j_3), X(j_4)\}$$

$$= \frac{1}{n^2}\left(\sum_{B\cup D} + \sum_C - \sum_D\right)\operatorname{Cum}\{X(j_1), X(j_2), X(j_3), X(j_4)\}.$$

The Brillinger condition (3) is applied as follows:

$$\begin{aligned}\frac{1}{n^2}\sum_{B\cup D}\operatorname{Cum}\{X(j_1), X(j_2), X(j_3), X(j_4)\} &= \frac{l-k}{n^2}\sum_{|\nu_i|\le\nu_0}|C_4(\nu_1,\nu_2,\nu_3)| \\ &\le \frac{l-k}{n^2}L_3^3\left(1-\frac{1}{\nu_0}\right)^3 \\ &< \left(\frac{l}{n}-\frac{k}{n}\right)^2 K_B.\end{aligned}$$

Also

$$\begin{aligned}\frac{1}{n^2}\sum_C|\operatorname{Cum}\{X(j_1), X(j_2), X(j_3), X(j_4)\}| &\le \sum_{\nu_0<|\nu_i|<l-k}|C_4(\nu_1,\nu_2,\nu_3)| \\ &\le \sum_{\nu_0<|\nu_i|} C_4(\nu_1,\nu_2,\nu_3) \\ &\le \frac{l-k}{n^2}\frac{L_4^3}{\nu_0^3} \\ &< \left(\frac{l}{n}-\frac{k}{n}\right)^2\frac{K_C}{\nu_0^3}.\end{aligned}$$

Furthermore,

$$\frac{1}{n^2}\sum_D|\operatorname{Cum}\{X(j_1), X(j_2), X(j_3), X(j_4)\}| < \left(\frac{l}{n}-\frac{k}{n}\right)^2\frac{\nu_0 K_D}{n}.$$

If we combine these inequalities for the fourth order cumulants with the result for the covariance function, we can then choose C independent of n such that

$$E\left(\frac{1}{\sqrt{n}}S_l - \frac{1}{\sqrt{n}}S_k\right)^4 < C\left(\frac{k}{n}-\frac{l}{n}\right)^2.$$

This, along with the fact that $S_0 \equiv 0$, implies tightness of the measures P_{X_n}, $n = 1, 2, \ldots$.

If the time series is assumed to be Gaussian, the fourth order moments can be defined in terms of the covariance function. Then tightness can be demonstrated for weakly stationary time series satisfying (1) and (2).

We denote by $\{B(t),\ t \in [0, 1]\}$ the standard Brownian motion process with continuous sample paths. Such a process is Gaussian with $B(0) = E\{B(t)\} \equiv 0$ and $K(s, t) = E\{B(s)B(t)\} = \min(s, t)$. The measure on $C[0, 1]$ corresponding to this process is Wiener measure denoted by W. If we let $B_X(t) = \{2\pi f(0)\}^{1/2}B(t)$ and let W_X be the corresponding measure on $C[0, 1]$ then we have the following result, an extension of a result of Johnson and Bagshaw (1974) which applies to ARMA processes.

THEOREM 2. *Under assumptions* (2) *and* (3),

$$P_{X_n} \Rightarrow W_X.$$

PROOF. Theorem 1 assures us that the finite dimensional distributions of P_{X_n} converge to those of W_X, and (9) implies that the sequence P_{X_n}, $n = 1, 2, \ldots$, is tight. The proof is completed by applying Theorem 12.3 of Billingsley (1968). □

Another way of expressing the conclusion of Theorem 2 is to say that the partial sum sequences $\theta_{X_n}(t)$ given by (7) converge weakly to Brownian motion; that is,

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

The conclusion of Theorem 2 holds under assumptions (1) and (2) if it is assumed additionally that the error variables are Gaussian.

4. The regression residual process for stationary error structure. We now consider the sequence of partial sums of regression residuals when the error process is a stationary time series. If the regressor functions are continuously differentiable on $[0, 1]$ and are linearly independent, then the bilinear form $g(s, t)$ is well defined. Then a sequence of functions $hg_n(\cdot)$, $n = 1, 2, \ldots$, on $C[0, 1]$ into itself may be defined as follows:

$$hg_n\{\theta_{X_n}(t)\} = \theta_{X_n}(t) - \mathbf{e}'_{n,nt}\mathbf{A}_n(\mathbf{A}'_n\mathbf{A}_n)^{-1} \times \left[\theta_{X_n}(1)\mathbf{g}(1) - \sum_{j=1}^{n} \{\mathbf{g}(j/n) - \mathbf{g}((j-1)/n)\} \times \theta_{X_n}((j-1)/n)\right].$$

Abel's partial summation formula implies that $hg_n\{\theta_{X_n(t)}\} \equiv \theta_{gX_n}(t)$; recall that $\theta_{gX_n}(t)$ is defined by (6). If P_{gX_n}, $n = 1, 2, \ldots$, denotes the set of measures

generated in $C[0, 1]$ by $hg_n\{\theta_{X_n}(t)\}$, then in the notation of Theorem 5.5 of Billingsley (1968), $P_{gX_n} = P_{X_n}(hg_n)^{-1}$.

Also define $hg(\cdot)$, a function from $C[0, 1]$ into itself, as follows:

$$\begin{aligned} hg\{B_X(t)\} &= B_{gX}(t) \\ &= B_X(t) - B_X(1)\int_0^t g(y, 1)\, dy \\ &\quad + \int_0^1 B_X(y)\left\{\int_0^t \frac{\partial}{\partial y} g(y, z)\, dz\right\} dy. \end{aligned} \tag{10}$$

This process has $B_{gX}(0) = E\{B_{gX}(t)\} \equiv 0$ and covariance kernel

$$\begin{aligned} K_{gX}(s, t) &= E\{B_{gX}(s) B_{gX}(t)\} \\ &= 2\pi f(0)\left\{\min(s, t) - \int_0^s \int_0^t g(y, z)\, dy\, dz\right\}. \end{aligned} \tag{11}$$

If P_{gX} is the measure generated in $C[0, 1]$ by $hg\{B_X(\cdot)\}$, then, again in the notation of Theorem 5.5 of Billingsley (1968), $P_{gX} = W_X(hg)^{-1}$. We now state the theorem which characterizes the limit of the sequences of partial sums of regression residuals when the error process is a stationary time series.

THEOREM 3. *Assume conditions* (2) *and* (3). *Further, assume* $g_r(t)$, $r = 0, \ldots, p$, *are continuously differentiable and linearly independent. Then*

$$P_{gX_n} \quad \Rightarrow \quad P_{gX}.$$

PROOF. The functions $hg_n(\cdot)$, $n = 1, 2, \ldots$, and $hg(\cdot)$ are continuous in the uniform topology on $C[0, 1]$. Furthermore, since $\lim_{n\to\infty} n^{-1}\mathbf{e}'_{n,nt}\mathbf{A}_n = \int_0^t \mathbf{g}'(s)\, ds$ and $\lim_{n\to\infty} n(\mathbf{A}'_n\mathbf{A}_n)^{-1} = \mathbf{G}^{-1}$, we see that $hg_n(\cdot)$ converges to $hg(\cdot)$. The proof follows by application of Theorem 5.5 of Billingsley (1968). □

Another way of expressing the conclusion of Theorem 3 is to say that the residual processes $\theta_{gX_n}(t)$ given by (6) converge weakly to limit processes $B_{gX}(t)$ given by (10); that is,

$$\theta_{gX_n}(t) \quad \Rightarrow \quad B_{gX}(t).$$

The conclusion of Theorem 3 holds under conditions (1) and (2) together with the additional assumption that the error process is Gaussian.

Results similar to (10) and (11) can be demonstrated for nonlinear regression. For such models the bilinear form $g(s, t)$ becomes a function also of the parameters. The additional assumptions required for the nonlinear case are provided by MacNeill and Jandhyala (1985).

The time of observation need not be restricted to $[0, 1]$ nor must sampling be equispaced. Suppose the total sampling period is $[0, T]$ and the rate of sampling is described by a nonconstant positive function $\{s(t),\ t \in [0, T]\}$. Then with $S(t) = \int_0^t s(x)\, dx / \int_0^T s(x)\, dx$, the limit process $B_{gXS}(\cdot)$ may be

related to that of Theorem 3 by the relation

$$B_{gXS}(t) = B_{gX}\{S(t)\}.$$

5. Examples of residual processes. The first example we consider is the case of fitting an harmonic polynomial of degree p to a set of data. Let $g_k(t) = \cos 2\pi kt$, $k = 0, 1, \ldots, p$, and $g_{p+k}(t) = \sin 2\pi kt$, $k = 1, 2, \ldots, p$. Furthermore let the noise process $X(n)$, $n = 0, \pm 1, \ldots$, be an autoregressive-moving average process defined by the equations

$$\sum_{j=0}^{r_1} b_j X(n-j) = \sum_{i=0}^{r_2} a_i \varepsilon(n-i), \qquad n = 0, \pm 1, \ldots, \tag{12}$$

where $\varepsilon(n)$, $n = 0, \pm 1, \ldots$, is normal white noise with zero mean and variance $\sigma^2 < \infty$, and where $A(z) = \sum_{i=0}^{r_2} a_i z^i$ and $B(z) = \sum_{i=0}^{r_1} b_i z^i$, $a_0 = b_0 = 1$, are functions of the complex variable z with zeros bounded away from the unit disc. The spectral density of the noise process is

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|A(e^{-i\lambda})|^2}{|B(e^{-i\lambda})|^2}, \qquad \lambda \in [-\pi, \pi].$$

Then (10) implies that the asymptotic residual process is

$$B_{gX}(t) = \sigma \frac{\sum_{j=0}^{r_2} a_j}{\sum_{j=0}^{r_1} b_j} \left[B(t) - B(1)\left\{ t + \sum_{k=1}^{p} (\pi k)^{-1} \sin 2\pi kt \right\} + 2\int_0^1 B(s) \left[\sum_{k=1}^{p} \{\cos 2\pi ks - \cos 2\pi k(t-s)\} \right] ds \right].$$

Equation (11) implies that the covariance kernel of the residual process is

$$K_{gX}(s,t) = \sigma^2 \frac{\left(\sum_{j=0}^{r_2} a_j\right)^2}{\left(\sum_{j=0}^{r_1} b_j\right)^2} \left[\min(s,t) - st - \sum_{j=1}^{p} \frac{1}{2\pi^2 j^2} \{(1 - \cos 2\pi jt)(1 - \cos 2\pi js) + \sin 2\pi js \sin 2\pi jt\} \right].$$

As another example we consider the case of fitting a polynomial of degree p to a set of data. Then $g_k(t) = t^k$, $k = 0, 1, \ldots, p$. Let the noise process $X(n)$, $n = 0, \pm 1, \ldots$, be a multiplicative seasonal moving average process (with 12 seasons) defined by the equation,

$$X(n) = 1 + a_1 \varepsilon(n) + a_{12} \varepsilon(n-12) + a_1 a_{12} \varepsilon(n-13),$$

where $\varepsilon(n)$, $n = 0, \pm 1, \ldots$, is as in the previous example and the parameters

are such that $|a_1| < 1$ and $|a_{12}| < 1$. The spectral density is

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|(1 + a_1 e^{-i\lambda})(1 + a_{12}e^{-12i\lambda})\right|^2.$$

Then (10) implies [see MacNeill (1978b)] that the asymptotic residual process is

$$B_{gX}(t) = \sigma(1 + a_1)(1 + a_{12})$$

$$\times\left[B(t) - \sum_{m=0}^{p}(2m+1)\left[B(1)g_m^*(1)\int_0^t g_m^*(s)\,ds - \left\{\int_0^t g_m^*(s)\,ds\right\}\left\{\int_0^1 B(s)\frac{\partial}{\partial s}g_m^*(s)\,ds\right\}\right]\right], \tag{13}$$

where

$$g_m^*(t) = \sum_{q=0}^{[m/2]}(-1)^q \frac{\binom{2m}{m,q,q,m-2q}}{2^{4q}\binom{m-\frac{1}{2}}{q}}\left(t - \frac{1}{2}\right)^{m-2q}.$$

Formula (11) implies that

$$K_{gX}(s,t) = \sigma^2(1 + a_1)^2(1 + a_{12})^2 \times\left[\min(s,t) - \sum_{m=0}^{p}(2m+1)\left\{\int_0^s g_m^*(x)\,dx\right\}\left\{\int_0^t g_m^*(y)\,dy\right\}\right].$$

6. Adjustments to account for the effect of serial correlation. In this section we indicate how the properties of the residual processes previously discussed can be used to adjust large sample distributional results for the class of statistics which are defined in terms of partial sums of residuals so as to account for serially correlated errors.

We let $F(\cdot)$ be a continuous functional defined on $C[a,b]$, the space of continuous functions on the interval $[a,b]$. Furthermore, we assume $F(\cdot)$ to be homogeneous of degree d; that is, if $f \in C[a,b]$ and k is a constant, then

$$F(kf) = k^d F(f).$$

Also we let $F_n(\cdot)$, $n = 1, 2, \ldots$, be a sequence of continuous functionals defined on $C[a,b]$ such that $F_n(\cdot) \to F(\cdot)$. Then if f_n, $n = 1, 2, \ldots$, and f are elements of $C[a,b]$ and if $f_n \to f$ uniformly on $[a,b]$, we have

$$k^d F_n(f_n) \to F(kf) = k^d F(f). \tag{14}$$

Hence, if we consider θ_{gX_n} defined by (6), $B_{gX}(\cdot)$ defined by (10), $B_g(\cdot)$ defined by $\{2\pi f(0)\}^{-1/2}B_{gX}(\cdot)$, and if $\tilde{R}_n(0)$ is a consistent estimator of $R(0) = \int_{-\pi}^{\pi} f(\lambda)\,d\lambda$, then a consequence of Donsker's theorem is the following result.

THEOREM 4. *If F_n, $n = 1, 2, \ldots$, and F are continuous functionals on $C[a, b]$ satisfying* (14), *then for $\alpha \in (0, 1)$*

$$(15) \qquad P\left[F_n\{\theta_{gX_n}(\cdot)\} > z_\alpha\{2\pi f(0)\}^{d/2}\right] \to P\left[F\{B_g(\cdot)\} > z_\alpha\right] = \alpha$$

and

$$(16) \quad P\left[\frac{F_n\{\theta_{gX_n}(\cdot)\}}{\{\tilde{R}_n(0)\}^{d/2}} > z_\alpha\left\{\frac{2\pi f(0)}{\int_{-\pi}^{\pi} f(\lambda)\,d\lambda}\right\}^{d/2}\right] \to P\left[F\{B_g(\cdot)\} > z_\alpha\right] = \alpha.$$

The implication of (15) is that the large sample quantiles of $F_n\{\theta_{gX_n}(\cdot)\}$ can be obtained from those of $F\{B_g(\cdot)\}$ simply by multiplying by $\{2\pi f(0)\}^{d/2}$. In the more likely event of normalization with a consistent estimator of the standard deviation, corresponding results are given in (16). Hence if distribution theory is available for the case of white noise error structure, then (15) and (16) give simple precise large sample adjustments to account for serial correlation in the noise process.

We have assumed that the functional $F(\cdot)$ is continuous on $C[a, b]$. For certain statistics it is possible to enlarge this class to include useful functionals that are not continuous; see MacNeill (1978a).

The availability of distribution theory for change-point statistics is discussed below, and adjustments for common noise models are discussed in the next section.

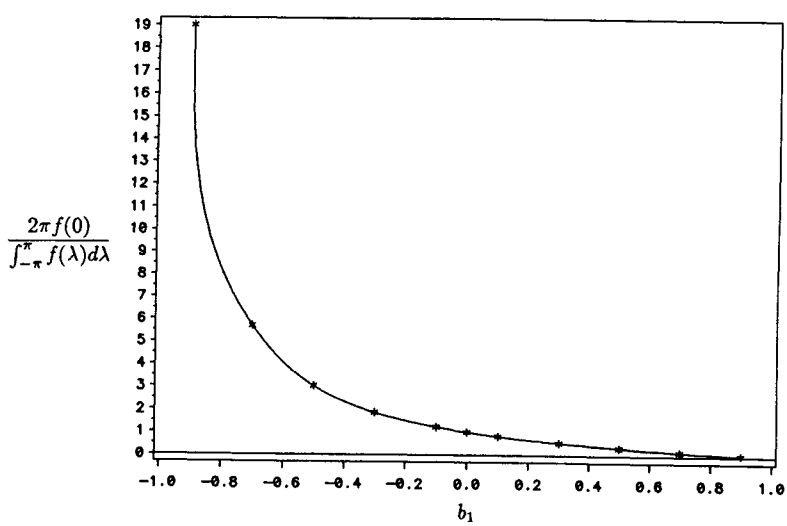


FIG. 2. *Plot of $\{2\pi f(0)\}/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda$ for AR(1) processes, $X(t) + b_1 X(t-1) = \varepsilon(t)$.*

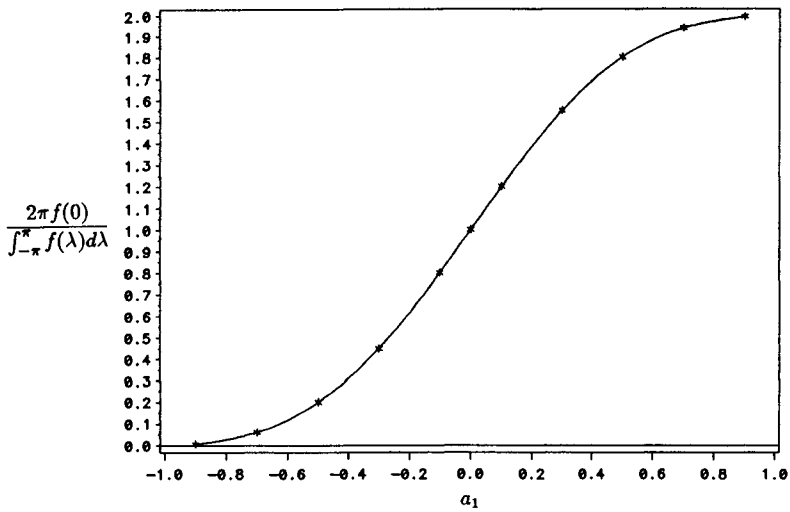
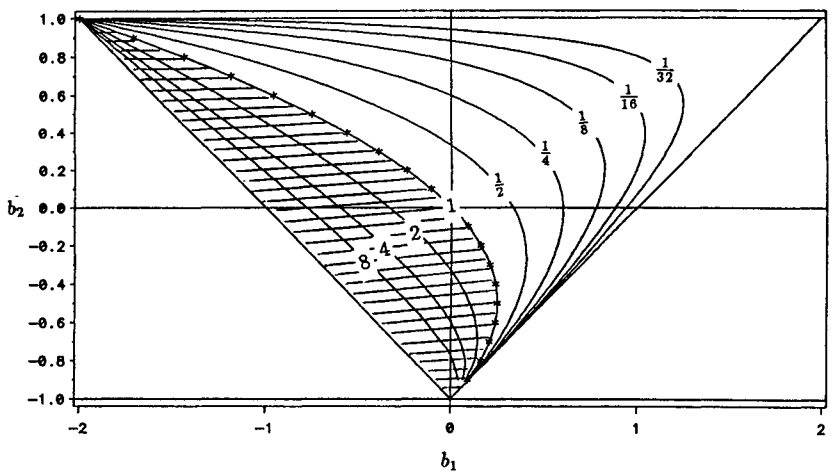
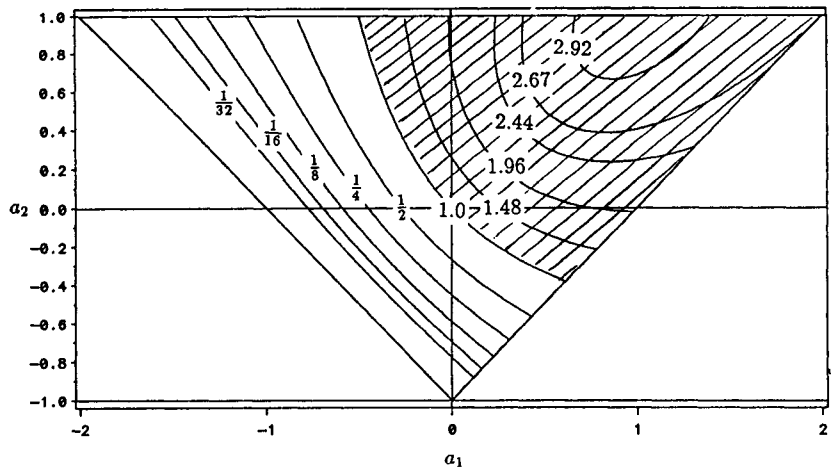


FIG. 3. *Plot of* $\{2\pi f(0)\}/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda$ *for MA*(1) *processes,* $X(t) = \varepsilon(t) + a_1\varepsilon(t-1)$.



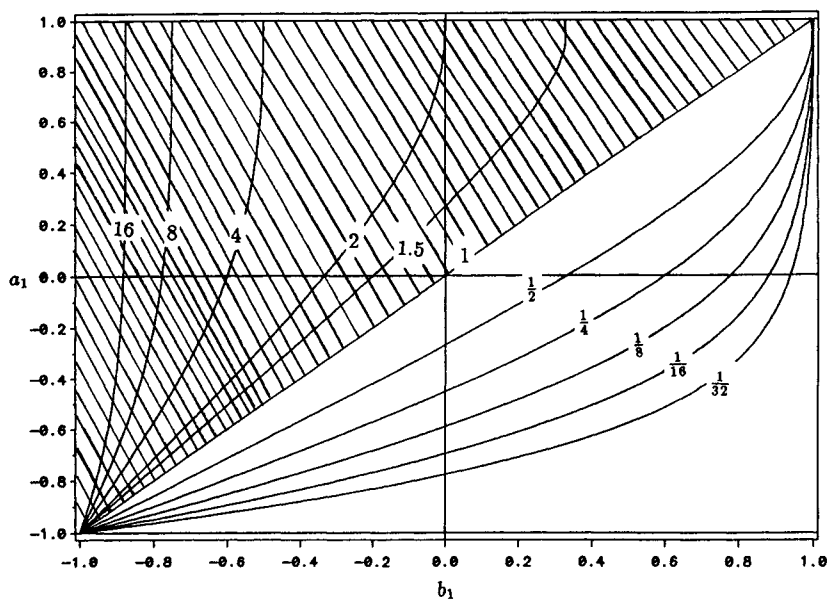
Parameter region where the zero frequency spectrum ordinate exceeds the average spectral density.

FIG. 4. *Plot of constant contours of* $\{2\pi f(0)\}/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda$ *for AR*(2) *processes,* $X(t) + b_1X(t-1) + b_2X(t-2) = \varepsilon(t)$.



Parameter region where the zero frequency spectrum ordinate exceeds the average spectral density.

FIG. 5. *Plot of constant contours of* $\{2\pi f(0)\}/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda$ *for MA*(2) *processes,* $X(t) = \varepsilon(t) + a_1\varepsilon(t-1) + a_2\varepsilon(t-2)$.



Parameter region where the zero frequency spectrum ordinate exceeds the average spectral density.

FIG. 6. *Plot of constant contours of* $\{2\pi f(0)\}/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda$ *for ARMA*(1, 1) *processes,* $X(t) + b_1X(t-1) = \varepsilon(t) + a_1\varepsilon(t-1)$.

7. Adjustments for common noise models to account for the effect of serial correlation. To give some insight into the nature of the adjustments required by (16) to account for serial correlation we have considered the following common models for the noise process: AR(1), AR(2), MA(1), MA(2) and ARMA(1, 1), where the ARMA(p, q) process is defined by (12).

The adjustment $2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda$ is plotted over the stationarity/invertibility region for the AR(1) and MA(1) processes in Figures 2 and 3. As can be noted from the figures the adjustment required to account for serial correlation in the noise process can be very significant if the parameters approach the edge of the stationarity/invertibility region. The contour curves for the adjustment factor are plotted for AR(2), MA(2) and ARMA(1, 1) in Figures 4, 5 and 6, respectively. The hatched areas of the stationarity/invertibility regions indicate where the adjustment is greater than 1. Again it can be noted that as the parameters approach the edges of the stationarity/invertibility regions the effect of the adjustments becomes large.

8. Effect of serial correlation on change detection statistics. We first consider Cramér–von Mises type statistics for testing for parameter change. For the case of white noise error structure with $\sigma^2 < \infty$, a statistic for detecting change at unknown time in regression parameters is shown by Jandhyala and MacNeill (1991) to be

$$Q_n(P) = \sum_{m=1}^{n-1} P_m \left\{ \frac{1}{\sigma\sqrt{n}} \sum_{k=m+1}^{n} \left(Y_n(k) - \hat{Y}_n(k)\right) \right\}^2. \tag{17}$$

Other Cramér–von Mises type statistics are derived by MacNeill (1982) and Jandhyala and MacNeill (1991) to deal with variations on the change-point theme. In (17), P_m, $m = 1, \ldots, n-1$, is a prior distribution on the change-point which, if the prior is uninformative, is $P_m = (n-1)^{-1}$. To make the statistic both operational and effective it is necessary to estimate σ^2 with an estimator that is consistent under both null and alternative hypotheses; in the event of polynomial regression these estimators may be based on several differences of the data.

Since the statistic in (17) is based on weighted "reversed" partial sums of residuals, large sample distribution theory may be based on limit processes of partial sums. If

$$P_{n-m} = \int_{(2m-1)/2n}^{(2m+1)/2n} \Psi(t)\,dt,$$

where $\Psi(t)$, $t \in [0, 1]$ is a nonnegative weight function such that $\int_0^1 t(1-t)\Psi(t)\,dt < \infty$ then $Q_n(P)$ converges in distribution as follows:

$$Q_n(P) \xrightarrow[\mathscr{D}]{} \int_0^1 \Psi(t)\, B_{gX}^2(t)\,dt.$$

Now assume the error process is not white noise and $R(0)$ is estimated consistently by $\tilde{R}(0)$ which is used in place of σ^2 in (17). Then, since the

statistic $Q_n(P)$ is a homogeneous functional of degree $d = 2$ on the normalized partial sum sequence, we have from (16) and the subsequent remarks regarding noncontinuous functionals that

$$Q_n(P) \xrightarrow[\mathscr{D}]{} \frac{2\pi f(0)}{\int_{-\pi}^{\pi} f(\lambda)\, d\lambda} \int_0^1 \Psi(t)\, B_g^2(t)\, dt. \tag{18}$$

The stochastic integral in (18) is defined in terms only of the weight function, Brownian motion and the regression part of model (5). The distribution of the stochastic integral in (18) has been tabulated for a variety of weight functions and regression models: polynomial regression with uniform prior by MacNeill (1978a); harmonic regression with uniform prior by Jandhyala and MacNeill (1989); Brownian motion with $\Psi(t) = at^k$, $k > -2$, by MacNeill (1974); Brownian motion, Brownian bridge and generalized Brownian bridges with Beta priors by Tang and MacNeill (1992); and the Brownian bridge process with $\Psi(t) = \{t(1-t)\}^{-1}$ by Anderson and Darling (1952).

Expression (18) indicates that the large sample effects of serial correlation on $Q_n(P)$ can be adjusted for precisely by multiplying the quantiles of distributions for the white noise case by $\{2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\, d\lambda\}$. To make this adjustment operational one will generally need to estimate $f(\lambda)$. This can be managed either by smoothing the periodogram of the residuals or by fitting a finite parameter scheme to the residuals. Care must be taken in estimating $f(0)$ to minimize regression induced biases.

In the case of white noise errors, $\{2\pi f(0)\}R(0)^{-1} = 1$ and the tabulated distribution can be used directly. If the noise process is an autoregression of order 1, that is, $X(t) + b_1 X(t-1) = \varepsilon(t)$, then

$$\frac{2\pi f(0)}{\int_{-\pi}^{\pi} f(\lambda)\, d\lambda} = \frac{2\pi\left(\sigma^2/\left(2\pi(1+b_1)^2\right)\right)}{\sigma^2/\left(1-b_1^2\right)} = \frac{1-b_1}{1+b_1}.$$

Hence quantiles of the tabulated distribution based on white noise errors must be multiplied by $(1-b_1)/(1+b_1)$ in order to obtain corresponding quantiles for the statistic with first order autoregressive errors. See Figure 2 for a plot of this adjustment factor. If one ignores the effect of serial correlation on $Q_n(P)$ and uses the tables based on the white noise error process, then, when b_1 is positive, one will compare this test statistic with quantiles that are too large by a factor of $[(1-b_1)/(1+b_1)]^{-1}$, and hence one will have reduced the power of the test; if b_1 is close to 1 then the power will have been greatly reduced. On the other hand, when b_1 is negative one will compare the test statistic with quantiles that are too small, again by the factor $[(1-b_1)/(1+b_1)]^{-1}$, and hence one will increase the probability of a type 1 error; if b_1 is close to -1 then this level of significance will be greatly increased from the putative level.

More generally, the factor $\{2\pi f(0)\}/\int_{-\pi}^{\pi} f(\lambda)\, d\lambda$ in (18) indicates that, if the zero frequency ordinate of the spectrum exceeds the average spectral density of the error process then ignoring serial correlation will result in increased probability of committing a type 1 error, and if the zero frequency ordinate is exceeded by the average spectral density then ignoring serial

correlation in the error process will diminish the power of the test. The more pronounced is the imbalance between low and high frequency spectral mass, the more pronounced are the effects upon the power and the size of the test. These results correspond to those of Johnson and Bagshaw (1974) who emphasized the nonrobustness of CUSUM tests to departures from independence.

As can be noted from the figures in Section 7 the effects of serial correlation on change-point statistics increase as the boundary separating stationarity from nonstationarity is approached by the parameters of the noise process. In the event the boundary is reached, the correction method suggested above breaks down since nonstationary noise and parameter changes are confounded in the change-point problem. While considering a related problem Jandhyala and MacNeill (1992) have shown a duality between testing for the constancy of regression coefficients under random walk and change-point alternatives. The case of random walk alternatives has been explored by Nabeya and Tanaka (1988).

A Bayes-type statistic for one-sided testing of a change in the mean was derived by Chernoff and Zacks (1964) in a paper that developed the methodology used subsequently to derive other Bayes-type change-point statistics. If $Y(i)$, $i = 1, \ldots, n$, is a sequence of independent variables with $\sigma^2 < \infty$, then the statistic is

$$T_n = \hat{\sigma}^{-1} \sum_{k=1}^{n} (k-1)\big(Y(k) - \bar{Y}\big) = \hat{\sigma}^{-1} \sum_{k=1}^{n-1} \left\{ \sum_{j=k+1}^{n} \big(Y(j) - \bar{Y}\big) \right\},$$

where $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(Y(i) - \bar{Y})^2$. If the variables are independently and normally distributed with known variance replacing $\hat{\sigma}^2$, then exact distributional results for T_n are easily computed under both null and alternative hypotheses. Since T_n is homogeneous of degree $d = 1$ in the partial sum sequence, the adjustment to account for serial correlation is $\{2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\, d\lambda\}^{1/2}$.

We now consider change-point tests based on likelihood methods. If $Y(i)$, $i = 1, \ldots, n$, is a sequence of independent normal variables with variance $\sigma^2 < \infty$, then for testing for a shift in mean, Hawkins (1977) derived a likelihood ratio test statistic equivalent to

$$V_n = \frac{\max_{1 \le k \le n-1} |T_k|}{\{n\hat{\sigma}_n^2\}^{1/2}},$$

where $T_k = \{(k/n)S_n - S_k\}/\{k(1 - k/n)\}^{1/2}$, $S_k = \sum_{i=1}^{k} Y(i)$ and $\hat{\sigma}_n^2$ is the maximum likelihood estimator of the variance. Worsley (1979) provided selected quantiles for V_n partly based on Bonferroni inequalities and partly based on simulations. The distribution of V_n does not converge to a fixed limit, but increases with n. However, it is possible to define a restricted likelihood ratio statistic based on $|T_k|$ maximized over $nt_0 \le k \le nt_1$, where t_0 and t_1 are bounded away from zero and one, respectively. James, James and Siegmund (1987) provided distributional results under both null and alternative hypotheses, and pointed out that the null restricted likelihood ratio statistic was asymptotically distributed as $\sup_{0 < t_0 \le t \le t_1 < 1} |B_0(t)|/\{t(1-t)\}^{1/2}$, where $B_0(\cdot)$

is the Brownian bridge. Since this statistic is a functional that is homogeneous of degree $d = 1$ in the partial sum sequence normalized by an estimate of the standard deviation, and since the large sample approximation is similarly homogeneous in the Brownian bridge, (16) indicates that the large sample adjustment required to account for serial correlation is $\{2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda\}^{1/2}$. Figures 2–6 may be used for these statistics by taking square roots of the ordinates. The effects of ignoring serial correlation on the power and size of the restricted likelihood ratio test are similar to those discussed above for Cramér–von Mises type tests, but do not apply to the likelihood ratio test which does not converge to a fixed limit.

Likelihood methods have been used by Quandt (1960), Worsley (1983) and Kim and Siegmund (1989) to derive tests for parameter change in simple linear regression with white noise error structure. The restricted large sample approximation to the likelihood ratio statistic is homogeneous of degree $d = 1$ in the normalized residuals. Hence Theorem 4 indicates that the adjustment to the quantiles provided by the distributional results of Kim and Siegmund (1989) to account for serial correlation is again $\{2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda\}^{1/2}$.

The CUSUM statistic proposed by Page (1954) for detecting one-sided changes in mean level is

$$W_n^+ = \max_{1\le k\le n}\left\{S_k - \min_{1\le j\le k} S_j\right\}.$$

The two-sided version of this statistic as proposed by Nadler and Robbins (1971) is

$$W_n = \max_{1\le k\le n} S_k - \min_{1\le k\le n} S_k.$$

The one-sided statistic normalized by the standard deviation of the observations has the same large sample distribution under the null hypothesis as does $\sup_{0\le t\le 1} B(t)$, where $B(\cdot)$ is Brownian motion. The normalized two-sided statistic has the same distribution as the range of Brownian motion. This distribution was derived under the null hypothesis by Feller (1951) and under certain contiguous alternatives by Nadler and Robbins (1971). Since the statistics are homogeneous of degree $d = 1$ in the partial sum sequence, Theorem 4 indicates the large sample adjustment required to account for serial correlation is again $\{2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda\}^{1/2}$.

Brown, Durbin and Evans (1975) proposed a technique based on recursive residuals for testing for the constancy of regression relationships over time. Under the null hypothesis of no change in the regression parameters and the assumption of white noise error structure, these residuals have zero mean, constant variance and are uncorrelated. A test was proposed based on CUSUM's of these residuals normalized by the usual estimator of the standard deviation. A recipe was given for determining a pair of lines such that the probability of the CUSUM process crossing one or both of the lines is the significance level of the test. The location of these lines is characterized by a single tabulated value defined by the significance level. Since the test is a homogeneous functional of degree $d = 1$ on the recursive residuals and since the CUSUM process con-

verges weakly to Brownian motion, (16) indicates the large sample adjustment required to account for serial correlation in the noise process is to adjust this parameter by multiplying by $\{2\pi f(0)/\int_{-\pi}^{\pi} f(\lambda)\,d\lambda\}^{1/2}$. This corresponds to the adjustment suggested by Cox (1975).

9. Applicability of large sample serial correlation adjustments to finite sample statistics. In deriving the adjustments to change-detection statistics that are required to account for serial correlation in the noise process, we have neglected terms which become small when the sample size becomes large. In this section we present the results of a simulation study which give some indication of the circumstances to which the asymptotic results apply and of those to which they should be applied with some caution. We considered simple linear regression with AR(1) error structure and applied the statistic $Q_n(P)$ in (17) with an uniformative prior on the change-point. We used a simulation of 1000 replications to estimate various parameters and quantiles of the distribution of the change point statistic. The estimates were obtained for time series lengths of 100, 200, 400, 800 and 1000, and for selected values of the AR(1) parameter ranging from -0.9 to 0.9. Table 1 presents the results of the simulation.

In broad terms the study suggests the following:

1. As expected in finite sample studies, the larger the sample the better the conformance of estimated quantiles and parameter estimates to the asymptotic theoretical quantiles and parameters.
2. The closer the ARMA parameters are to the edge of the stationarity/invertibility region, the poorer the conformance of estimated quantiles and parameter estimates to the asymptotic theoretical quantiles and parameters.
3. Adjustment factors tend to be underestimated in small samples. It is well known that autoregression parameter estimates can have large biases for small samples.
4. In the case of predominance of low frequency spectral mass in the error process, which is the most common situation encountered in practice, the estimated adjustment for serial correlation to change-point statistics tends to produce conservative tests. The opposite is the case when high frequencies predominate.
5. As with all simulations, the effects of sample size on the distributions are felt most in the tails, with the 0.90 quantiles being better estimated than the 0.99 quantiles for all sample sizes.
6. In all cases the adjustments result in substantially more reliable tests than are obtained by ignoring the effects of serial correlation.

The specific conclusions suggested by this simulation study of AR(1) noise are as follows. The large sample theory can be applied with confidence to tests of size larger than 5% for time series lengths of 100 or more observations provided $|b_1| < 0.5$. By this we mean that the simulated quantile and the quantile obtained by multiplying the white noise quantile by the estimate of $\{2\pi f(0)\}/\{\int_{-\pi}^{\pi} f(\lambda)\,d\lambda\}$ do not differ by more than roughly 10%. If time series

TABLE 1

Simulation for selected parameters and quantiles of $Q_n(P)$: simple linear regression with AR(1) *noise, for selected values of the* AR *parameter b_1 and with an uniformative prior P on the change-point [compare items* (a), (b), (c) *and* (d)][1]

Parameter	Series size	$b_1 = 0.1$ (a)	$b_1 = 0.1$ (b)	$b_1 = 0.3$ (a)	$b_1 = 0.3$ (b)	$b_1 = 0.5$ (a)	$b_1 = 0.5$ (b)	$b_1 = 0.7$ (a)	$b_1 = 0.7$ (b)	$b_1 = 0.9$ (a)	$b_1 = 0.9$ (b)
$\bar{Q}_n(P)$	100	0.0553	0.0511	0.0399	0.0342	0.0257	0.0216	0.0151	0.0113	0.0071	0.0033
	200	0.0553	0.0532	0.0361	0.0351	0.0240	0.0218	0.0134	0.0116	0.0054	0.0035
	400	0.0539	0.0537	0.0364	0.0355	0.0228	0.0221	0.0125	0.0116	0.0045	0.0034
	800	0.0541	0.0541	0.0357	0.0357	0.0220	0.0221	0.0118	0.0117	0.0040	0.0035
	1000	0.0543	0.0543	0.0363	0.0358	0.0221	0.0222	0.0121	0.0117	0.0039	0.0035
[(d) 0.0667]	(c)	0.0545		0.0359		0.0222		0.0118		0.0035	
$Q_{n,0.90}(P)$	100	0.0988	0.0913	0.0707	0.0611	0.0450	0.0387	0.0252	0.0202	0.0110	0.0059
	200	0.0944	0.0950	0.0632	0.0628	0.0428	0.0390	0.0229	0.0207	0.0090	0.0062
	400	0.0945	0.0960	0.0624	0.0634	0.0409	0.0395	0.0226	0.0208	0.0078	0.0061
	800	0.0967	0.0966	0.0633	0.0638	0.0387	0.0395	0.0215	0.0208	0.0072	0.0062
	1000	0.0942	0.0971	0.0632	0.0640	0.0381	0.0396	0.0221	0.0209	0.0068	0.0063
[(d) 0.1192]	(c)	0.0975		0.0642		0.0397		0.0210		0.0063	
$Q_{n,0.95}(P)$	100	0.1183	0.1133	0.0887	0.0759	0.0552	0.0480	0.0310	0.0250	0.0137	0.0074
	200	0.1174	0.1179	0.0747	0.0779	0.0511	0.0484	0.0297	0.0256	0.0114	0.0077
	400	0.1140	0.1191	0.0832	0.0787	0.0495	0.0490	0.0269	0.0257	0.0093	0.0076
	800	0.1158	0.1199	0.0772	0.0792	0.0465	0.0490	0.0268	0.0259	0.0088	0.0077
	1000	0.1176	0.1204	0.0798	0.0793	0.0475	0.0492	0.0263	0.0260	0.0084	0.0078
[(d) 0.1479]	(c)	0.1210		0.0796		0.0493		0.0261		0.0078	
$Q_{n,0.99}(P)$	100	0.1805	0.1667	0.1293	0.1117	0.0793	0.0706	0.0556	0.0368	0.0196	0.0109
	200	0.1668	0.1736	0.1080	0.1147	0.0753	0.0713	0.0444	0.0377	0.0175	0.0113
	400	0.1685	0.1753	0.1197	0.1158	0.0650	0.0721	0.0382	0.0379	0.0118	0.0112
	800	0.1682	0.1765	0.1064	0.1166	0.0626	0.0722	0.0376	0.0381	0.0124	0.0114
	1000	0.2014	0.1773	0.1105	0.1168	0.0720	0.0724	0.0389	0.0382	0.0123	0.0114
[(d) 0.2177]	(c)	0.1782		0.1172		0.0726		0.0384		0.0115	

lengths are 200 or more, the large sample theory can be applied to tests of size 5% or larger provided $|b_1| < 0.7$. If time series lengths are 1000 or more, the large sample theory can be applied to tests of size 5% or larger provided $|b_1| < 0.9$. In these cases the tests are conservative in the common situation when low frequency spectral mass predominates, that is, when $b_1 < 0$.

Johnson and Bagshaw's (1974) study of the effect of serial correlation on run length distribution indicates that when low frequency power predominates, the run length for CUSUM tests is substantially shortened, while the opposite holds for the case when high frequency power predominates. Their simulation only considers the case of $|b_1| \leq 0.5$ but it appears that a preponderance of high frequency power produces finite sample average run lengths which exceed those suggested by asymptotic theory while a preponderance of low frequency power produces finite sample average run lengths which are shorter. These results also confirm that as the imbalance between high and

TABLE 1 (*Continued*)

Parameter	Series size	$b_1 = -0.1$ (a)	$b_1 = -0.1$ (b)	$b_1 = -0.3$ (a)	$b_1 = -0.3$ (b)	$b_1 = -0.5$ (a)	$b_1 = -0.5$ (b)	$b_1 = -0.7$ (a)	$b_1 = -0.7$ (b)	$b_1 = -0.9$ (a)	$b_1 = -0.9$ (b)
$\overline{Q}_n(P)$	100	0.0804	0.0765	0.1128	0.1128	0.1647	0.1753	0.2720	0.3066	0.5019	0.7123
	200	0.0811	0.0790	0.1194	0.1197	0.1954	0.1886	0.3230	0.3448	0.7504	0.9324
	400	0.0801	0.0801	0.1224	0.1215	0.1986	0.1939	0.3439	0.3625	0.8851	1.0912
	800	0.0813	0.0811	0.1194	0.1229	0.1914	0.1975	0.3545	0.3692	1.1533	1.1787
	1000	0.0778	0.0807	0.1225	0.1227	0.1949	0.1974	0.3641	0.3721	1.1300	1.1942
[(d) 0.0667]	(c)	0.0815		0.1238		0.2000		0.3778		1.2540	
$Q_{n,0.90}(P)$	100	0.1416	0.1368	0.1985	0.2015	0.2916	0.3133	0.4944	0.5479	0.9334	1.2729
	200	0.1474	0.1412	0.2023	0.2138	0.3729	0.3370	0.5870	0.6162	1.3614	1.6662
	400	0.1467	0.1432	0.2243	0.2171	0.3534	0.3466	0.5843	0.6479	1.8028	1.9501
	800	0.1438	0.1449	0.2071	0.2196	0.3412	0.3529	0.6434	0.6599	2.0367	2.1066
	1000	0.1351	0.1442	0.2208	0.2192	0.3411	0.3527	0.6618	0.6650	2.0349	2.1341
[(d) 0.1192]	(c)	0.1457		0.2214		0.3577		0.6756		2.2652	
$Q_{n,0.95}(P)$	100	0.1777	0.1697	0.2471	0.2501	0.3489	0.3888	0.6148	0.6798	1.1119	1.5794
	200	0.1772	0.1752	0.2630	0.2653	0.4631	0.4181	0.7296	0.7646	1.6448	2.0674
	400	0.1803	0.1777	0.2854	0.2694	0.4327	0.4300	0.7243	0.8039	2.1961	2.4196
	800	0.1714	0.1797	0.2510	0.2725	0.4272	0.4379	0.7356	0.8187	2.5018	2.6137
	1000	0.016430.1789		0.2818	0.2720	0.4165	0.4377	0.7653	0.8251	2.5194	2.6479
[(d) 0.1479]	(c)	0.1808		0.2747		0.4437		0.8380		2.8099	
$Q_{n,0.99}(P)$	100	0.2543	0.2498	0.3442	0.3681	0.4983	0.5722	0.8467	1.0006	1.4671	2.3248
	200	0.2630	0.2579	0.3614	0.3905	0.6369	0.6155	1.0323	1.1255	2.3203	3.0431
	400	0.2483	0.2615	0.4329	0.3965	0.5888	0.6330	1.0507	1.1833	2.7253	3.5615
	800	0.2596	0.2646	0.3690	0.4011	0.6195	0.6446	1.0517	1.2051	3.6154	3.8473
	1000	0.2293	0.2634	0.4250	0.4004	0.6070	0.6442	1.1481	1.2145	3.3893	3.8975
[(d) 0.2177]	(c)	0.2661		0.4044		0.6532		1.2339		4.1372	

[1](a) Derived through simulation; (b) estimated by adjusting white noise values by $2\pi \hat{f}(0)/\int_{-\pi}^{\pi} \hat{f}(\lambda)\, d\lambda$; (c) theoretical value; (d) white noise value to be used if serial correlation is ignored [see Table 1 in Jandhyala and MacNeill (1991)].

low frequency power increases, the adequacy of asymptotic theory decreases.

Acknowledgment. The authors are grateful to the referees for providing helpful suggestions that resulted in improvements to the paper.

REFERENCES

ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

BAGSHAW, M. and JOHNSON, R. A. (1975). The effect of serial correlation on the performance of CUSUM tests. *Technometrics* **17** 73–80.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BRILLINGER, D. R. (1973). Estimation of the mean of a stationary time series by sampling. *J. Appl. Probab.* **10** 419–431.

BRILLINGER, D. R. (1989). Consistent detection of a monotonic trend superposed on a stationary time series. *Biometrika* **76** 23–30.

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192.

CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Ann. Math. Statist.* **35** 999–1018.

COX, D. R. (1975). Discussion of "Techniques for testing the constancy of regression relationships over time" by R. L. Brown, J. Durbin and J. M. Evans. *J. Roy. Statist. Soc. Ser. B* **37** 163–164.

EL-SHAARAWI, A. H. and ESTERBY, S. R. (1982). Inference about the point of change in a regression model with a stationary error process. In *Time Series Methods in Hydrosciences* (A. H. El-Shaarawi and S. R. Esterby, eds.) 55–67. North-Holland, Amsterdam.

FELLER, W. (1951). The asymptotic distribution of the range of sums of independent variables. *Ann. Math. Statist.* **22** 427–432.

GRENANDER, V. (1954). On the estimation of the regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25** 252–272.

HANNAN, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probab.* **10** 130–145.

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.* **72** 180–186.

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74** 70–83.

JANDHYALA, V. K. and MACNEILL, I. B. (1989). Residual partial sum limit processes for regression models with application to detecting parameter changes at unknown times. *Stochastic Process. Appl.* **33** 309–323.

JANDHYALA, V. K. and MACNEILL, I. B. (1991). Tests for parameter changes at unknown times in linear regression models. *J. Statist. Plann. Inference* **27** 291–316.

JANDHYALA, V. K. and MACNEILL, I. B. (1992). On testing for the constancy of regression coefficients under random walk and change-point alternatives. *Econometric Theory* **8** 501–507.

JOHNSON, R. A. and BAGSHAW, M. (1974). The effect of serial correlation on the performance of CUSUM tests. *Technometrics* **16** 103–112.

KIM, H.-J. and SEIGMUND, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika* **76** 409–423.

KULPERGER, R. J. (1985). On the residuals of autoregressive processes and polynomial regression. *Stochastic Process. Appl.* **21** 107–118.

MACNEILL, I. B. (1974). Tests for change of parameter at unknown times and distributions of some related functionals on Brownian motion. *Ann. Statist.* **2** 950–962.

MACNEILL, I. B. (1978a). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change in regression at unknown times. *Ann. Statist.* **6** 422–433.

MACNEILL, I. B. (1978b). Limit processes for sequences of partial sums of regression residuals. *Ann. Probab.* **6** 695–698.

MACNEILL, I. B. (1982). Detecting interventions at unknown times. In *Time Series Methods in Hydrosciences* (A. H. El-Shaarawi and S. R. Esterby, eds.) 16–26. North-Holland, Amsterdam.

MACNEILL, I. B. and JANDHYALA, V. K. (1985). The residual process for non-linear regression. *J. Appl. Probab.* **22** 957–963.

NABEYA, S. and TANAKA, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Statist.* **16** 218–235.

NADLER, J. and ROBBINS, N. D. (1971). Some characteristics of Page's two-sided procedure for detecting a change in a location parameter. *Ann. Math. Statist.* **42** 538–551.

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–116.

PICARD, D. (1985). Testing and estimating change-points in time series. *Adv. in Appl. Probab.* **17** 841–867.

QUANDT, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Assoc.* **55** 324–330.

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Internat. Statist. Rev.* **56** 31–48.
TANG, S. M. and MACNEILL, I. B. (1992). Monitoring statistics which have increased power over a reduced time range. *Environmental Monitoring and Assessment.* To appear.
WORSLEY, K. J. (1979). On the likelihood ratio test for a shift in loction of normal populations. *J. Amer. Statist. Assoc.* **74** 365–367.
WORSLEY, K. J. (1983). Testing for a two phase multiple regression. *Technometrics* **25** 35–42.
WORSLEY, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family variables. *Biometrika* **73** 91–104.

COMPUSEARCH MARKET
AND SOCIAL RESEARCH LTD.
330 FRONT STREET WEST
SUITE 110
TORONTO
CANADA M5V 3B7

DEPARTMENT OF STATISTICAL
AND ACTUARIAL SCIENCES
UNIVERSITY OF WESTERN ONTARIO
LONDON
CANADA N6A 5B7