Statistica Sinica **13**(2003), 827-852

ROBUST TESTS FOR INDEPENDENCE OF TWO TIME SERIES

Pierre Duchesne and Roch Roy

Université de Montréal

Abstract: This paper aims at developing a robust and omnibus procedure for checking the independence of two time series. Li and Hui (1994) proposed a robustified version of Haugh's (1976) classic portmanteau statistic which is based on a fixed number of lagged residual cross-correlations. Hong (1996a) introduced a class of consistent test statistics which are weighted sums of residual cross-correlations. These tests provide a generalized Haugh's test statistic. Hong's tests are sensitive to outliers, since they are based on the usual cross-correlation function and least-squares estimators. Here, we introduce a robustified version of Hong's test statistics. We suppose that for each series, the true ARMA model is estimated by a $n^{1/2}$-consistent method. If outliers are suspected, robust estimators of the parameters are obtained and the new test statistics rely on the sample robust cross-correlation function introduced in Li and Hui (1994). Under the null hypothesis of independence, the new tests asymptotically follow a $N(0,1)$ distribution. Using the truncated uniform kernel, our test provides a generalized version of the robust test statistic of Li and Hui (1994). We also propose a robust procedure for checking independence at individual lags and a descriptive causality in mean analysis in the Granger sense is discussed. From simulation results, we find that Hong's and Haugh's tests can be severely affected by additive outliers in the time series. The new robust statistics and the test of Li and Hui have reasonable levels when outliers are present. However, using a kernel different from the truncated uniform kernel, our test statistics may be substantially more powerful than the test of Li and Hui. Finally, the proposed robust procedures are applied to a set of financial data.

Key words and phrases: ARMA model, causality in mean, coherency, independence, robust estimation, robust serial correlation.

1. Introduction

Several physical and economic phenomena can be described by time series; see among others Akaike and Kitagawa (1999) for physical applications and Judge, Hill, Griffiths, Lütkepohl and Lee (1985) for economic applications. When we have to deal with two time series, the question often arises of describing the interrelationships existing between them. In economics, for example, elucidation of causality relationships between time series may be very important in a prediction context. Before applying sophisticated methods for describing relationships between two time series, it is important to check whether they are independent (or serially uncorrelated in the non-Gaussian case) or not. If sufficiently powerful methods, both simple to apply and interpret, were available for checking the independence of two time series, more sophisticated multivariate analyses such as causality in mean analysis might become redundant. Haugh's (1976) paper is to our knowledge the first attempt at developing a procedure based on residual cross-correlations for checking the independence of two stationary ARMA series. He considered a portmanteau statistic given by

$$S_M = n \sum_{j=-M}^{M} r_{\hat{u}\hat{v}}^2(j), \tag{1}$$

where $r_{\hat{u}\hat{v}}(j) = \sum_{t=j+1}^{n} \hat{u}_t \hat{v}_{t-j} / (\sum_{t=1}^{n} \hat{u}_t^2 \sum_{t=1}^{n} \hat{v}_t^2)^{1/2}$ are the residual cross-correlations for $0 \leq j \leq n-1$, $r_{\hat{u}\hat{v}}(j) = r_{\hat{v}\hat{u}}(-j)$ for $1-n \leq j < 0$, and $\hat{u}_t$, $\hat{v}_t$, $t = 1, \ldots, n$, are the two residual series of length n, obtained by fitting univariate models to each of the series. The constant $M \leq n-1$ is a fixed integer and must be chosen a priori. The asymptotic distribution of S_M is chi-square under the null hypothesis of independence and the hypothesis is rejected for large values of the test statistic. The constant M cannot be too large in order that the asymptotic chi-square approximation be satisfactory. However, since S_M does not take into account the cross-correlations at lag j for $j > M$, it will not detect serial correlation at high lags and leads therefore to an inconsistent test procedure.

Haugh's procedure was extended in various directions. Koch and Yang (1986) introduced a modification of S_M that allows for a potential pattern in the residual cross-correlation function. El Himdi and Roy (1997) proposed a version of S_M for two stationary vector ARMA (VARMA) that was recently extended to partially non stationary (cointegrated) VARMA series by Pham, Roy and Cedras (2003). Hallin and Saidi (2001) have proposed a generalization of the Koch and Yang procedure for VARMA series.

Hong (1996a) proposed a consistent test which is a generalization of Haugh's statistic that takes into account all possible lags. The test statistic is a weighted sum of residual cross-correlations of the form

$$Q_n = \frac{n \sum_{j=1-n}^{n-1} k^2(j/m) r_{\hat{u}\hat{v}}^2(j) - M_n(k)}{[2V_n(k)]^{1/2}}, \tag{2}$$

where $M_n(k) = \sum_{j=1-n}^{n-1} (1 - |j|/n) k^2(j/m)$ and $V_n(k) = \sum_{j=2-n}^{n-2} (1 - |j|/n)(1 - (|j|+1)/n) k^4(j/m)$. The weighting depends on a kernel function k and a smoothing parameter m as in kernel-based spectral density estimation; this allows a flexible weighting. With the truncated uniform kernel, Q_n corresponds to a normalized version of Haugh's statistic. However several kernels lead to higher power. Indeed, the so-called Daniell kernel maximizes the asymptotic power of the test Q_n in a certain class of smooth kernels and under some local alternatives. Also, Hong (1996a) avoids the ARMA modeling by fitting autoregressions of appropriate high orders. Under the null hypothesis, the test statistic Q_n is asymptotically $N(0,1)$. The test is unilateral and rejects for the large values of Q_n.

In practice, outliers in time series can create serious problems. They can occur for various reasons, measurement errors or equipment failure, etc. (see, e.g., Martin and Yohai (1985), Hampel, Ronchetti, Rousseeuw and Stahel (1986) and Rousseeuw and Leroy (1987)). Haugh and Hong tests for independence are based on estimation methods that are sensitive to outliers. Furthermore, the usual cross-correlation function can be considerably affected by outliers. Robust estimation methods in ARMA models were studied in Bustos and Yohai (1986), and a robustified autocovariance function introduced. Robust estimation of VARMA models was considered in Ben, Martinez and Yohai (1999). Li and Hui (1994) proposed a robustified cross-correlation function between two time series and employed it to develop a robust version of Haugh's (1976) and McLeod's (1979) tests for checking independence. Their robust statistics asymptotically follow a chi-square distribution whose degrees of freedom depend on the autoregressive and moving-average orders. Another approach of the nonparametric type for checking the independence of two autoregressive time series based on autoregressive rank scores was studied by Hallin, Jurevcková, Picek and Zahaf (1999).

The main objective of this paper is to develop a robust version of Hong's statistic for checking the independence of two univariate stationary and invertible time series. We suppose that for each series, the true ARMA model is estimated by a $n^{1/2}$-consistent method. If outliers are suspected in the time series, robust estimators of the parameters are obtained using, for example, the residual autocovariances estimators (RA estimators) introduced by Bustos and Yohai (1986). A robustified version of Q_n is obtained by replacing the usual residual cross-correlation in (2) by the robust cross-correlation function introduced by Li and Hui (1994). The new test has an asymptotic normal distribution under the null hypothesis of independence. Using the truncated uniform kernel, we retrieve Li and Hui's test statistic. Using a result of Li and Hui (1994), we also describe a robust procedure for checking independence at individual lags. As emphasized by Box, Jenkins and Reinsel (1994) and others, it is important not only to look at the value of a global statistic but also to examine the values of the cross-correlations at individual lags. It may well happen that the procedure based on individual lags lead us to reject the null hypothesis whilst the global test does not reject. Also, if in a first step we reject with the global test, the tests at individual lags will identify the lags where there is a significant correlation.

The organization of the paper is as follows. In Section 2, we introduce the notations and concepts employed thereafter. Our main result is presented in Section 3, where a robustified version of Hong's (1996a) statistic is introduced. Invoking a general result in Li and Hui (1994), we describe in Section 4 a robust procedure for checking independence at individual lags and we present a robustified version of Haugh's (1976) statistic. A descriptive approach for causality in mean analysis in Granger's (1969) sense is also discussed. In Section 5, we describe the results of a small Monte Carlo experiment conducted to analyze the level and power of the robust and non-robust test statistics. It is found that the Hong and Haugh tests can be severely affected by additive outliers in the time series. As for Li and Hui's statistic, the level of the new tests is reasonably well controlled when outliers are present. However, using a kernel different from the truncated uniform kernel, our test can be substantially more powerful than the test of Li and Hui. The new procedures are applied to a set of financial data in Section 6.

2. Preliminaries

2.1. Assumptions on the process

Let $\{(X_t, Y_t), t \in \mathbb{Z}\}$ be a bivariate second-order stationary process. Without loss of generality, we can assume that $\{(X_t, Y_t)\}$ is centered at zero. We further suppose that $\{X_t\}$ and $\{Y_t\}$ are invertible ARMA processes, that is $\phi_1(B)X_t = \theta_1(B)u_t$, $\phi_2(B)Y_t = \theta_2(B)v_t$, where $\phi_h(B) = \sum_{i=0}^{p_h} \phi_{hi}B^i$ are the autoregressive (AR) polynomials, $\theta_h(B) = \sum_{i=0}^{q_h} \theta_{hi}B^i$ the moving-average (MA) polynomials with $\phi_{h0} = \theta_{h0} = 1$, $h = 1, 2$, and B is the backward shift operator. We denote $\boldsymbol{\phi}_h = (\phi_{h1}, \ldots, \phi_{hp_h})'$ and $\boldsymbol{\theta}_h = (\theta_{h1}, \ldots, \theta_{hq_h})'$, $h = 1, 2$.

The innovation processes $\{u_t\}$ and $\{v_t\}$ are strong white noise, that is the u_t's and the v_t's are two sequence of independent and identically distributed (i.i.d.) random variables with mean zero and variance σ_u^2 and σ_v^2, respectively. It is also supposed that the fourth-order moments of u_t and v_t exist and that the probability distributions of u_t and v_t are symmetric with respect to zero. The symmetry assumption is frequent in the robustness context, see for example Denby and Martin (1979), Bustos, Fraiman and Yohai (1984) and Bustos and Yohai (1986). M-estimation in AR(p) models under nonstandard conditions is studied in El Bantli and Hallin (2001). Note that in linear regression models with asymmetric errors, some M-estimators of the slope parameters are consistent whilst the Generalized M (GM) estimators may be inconsistent (Carroll and Welsh (1988)). Huber (1981, p.171) gives arguments suggesting that, in general, the bias of M-estimators should not be too important in many practical applications; see also Huber (1964, Section 7). An interesting discussion of the role of symmetry in robust statistics is presented in Hampel, Ronchetti, Rousseeuw and Stahel (1986, p.401).

2.2. Outliers in time series

In time series, Fox (1972) introduced innovation outliers (IO) and additive outliers (AO). That terminology is frequently used in robustness studies, see for example Martin and Yohai (1985), Rousseeuw and Leroy (1987) and Hampel et al. (1986). Here, we briefly describe these two types of outliers using probability models as it is often done in distributional studies.

Suppose that $\{X_t\}$ is an ARMA(p,q) process and that the innovation process $\{u_t\}$ has an heavy tail distribution F which is not far from the Gaussian distribution, as for example a contaminated normal, $F = (1-\epsilon)N(0,\sigma^2) + \epsilon N(0,\tau^2)$, where $\epsilon > 0$ is small, and $\tau \geq \sigma$. In this situation, the innovations u_t are from a $N(0,\sigma^2)$ with probability $1-\epsilon$, or from a $N(0,\tau^2)$ with a larger variance with probability ϵ. The latter innovations are considered as outliers. The important point with IO is that the ARMA(p,q) model is still the exact model for the observations. However, if an outlier occurs at t_0, then u_{t_0} will affect not only X_{t_0}, but many future observations. After a while, the effect disappears. Bustos and Yohai (1986) give several results showing that IO do not affect too seriously the least squares (LS) estimators of autoregressive and moving average parameters of an ARMA process.

We now discuss additive outliers. Suppose that the observations are obtained from the model $X_t = \tilde{X}_t + V_t$, where $\{\tilde{X}_t\}$ and $\{V_t\}$ are independent, $\{\tilde{X}_t\}$ is ARMA with innovations u_t that are Gaussian with a common variance σ^2, the V_t's are i.i.d. random variables whose distribution is $H = (1-\epsilon)\delta_0 + \epsilon G$, where δ_0 is the degenerate distribution at zero and G is an arbitrary distribution. Here the ARMA process itself is observed without error with probability $1-\epsilon$ and there is a probability ϵ that the process ARMA plus an error of distribution G be obtained. When AO appear, we have an imperfect ARMA process. LS estimators and even M-estimators are quite sensitive to AO.

Bustos and Yohai (1986) introduced RA estimators (for *Residual Autocovariances*) as an alternative with good behavior when AO are present. These estimators are obtained by solving a system of equations similar to the one leading to the LS estimators, except that the usual sample innovation autocorrelation function is replaced by a robustified version. Bustos and Yohai (1986) also introduced a class of truncated RA estimators (TRA). An heuristic argument shows that these estimators are qualitatively robust when a MA component is present in the ARMA model. The main result of this paper described in Section 3 only requires that the AR and MA estimators be $n^{1/2}$-consistent. If outliers are suspected, the estimators need also to be robust, since the LS estimators may be severely biased when AO are present. For example, GM estimators as well as RA and TRA estimators are $n^{1/2}$-consistent and robust. RA estimators with a Mallows type function are particularly convenient to compute since an iterative scheme is possible. For rigor and completeness, we state explicitly the consistency condition in Assumption A.

Assumption A. Let $\{X_t\}$ and $\{Y_t\}$ be ARMA processes. Let $\hat{\boldsymbol{\lambda}}_u = (\hat{\boldsymbol{\phi}}_1', \hat{\boldsymbol{\theta}}_1')'$ and $\hat{\boldsymbol{\nu}}_u = (\hat{\boldsymbol{\lambda}}_u', \hat{\sigma}_u)'$ be the estimators of the parameters $\boldsymbol{\lambda}_u = (\boldsymbol{\phi}_1', \boldsymbol{\theta}_1')'$ and $\boldsymbol{\nu}_u = (\boldsymbol{\lambda}_u', \sigma_u)'$, respectively. Similar notations hold for $\{Y_t\}$. We suppose that $\hat{\boldsymbol{\nu}}_u - \boldsymbol{\nu}_u = \mathbf{O}_p(n^{-1/2})$ and $\hat{\boldsymbol{\nu}}_v - \boldsymbol{\nu}_v = \mathbf{O}_p(n^{-1/2})$.

With two time series, the cross-correlation function (CCF) is often used to appreciate the existing dependency between the two processes. However, as with the usual autocorrelation function, it is sensitive to outliers. Li and Hui (1994) proposed the following robust CCF:

$$\gamma_{uv}(j;\eta) = \begin{cases} n^{-1} \sum_{t=j+1}^{n} \eta(u_t/\sigma_u, v_{t-j}/\sigma_v), \ j \geq 0, \\ n^{-1} \sum_{t=-j+1}^{n} \eta(u_{t+j}/\sigma_u, v_t/\sigma_v), \ j < 0, \end{cases} \tag{3}$$

where $\{u_t\}$ and $\{v_t\}$ are the two innovation processes, with variances σ_u^2 and σ_v^2 respectively. The function $\eta(\cdot,\cdot)$ is to be continuous and odd in each variable. When $\eta(u,v) = uv$, we retrieve the usual sample CCF, and we write $\gamma_{uv}(j;\eta) = r_{uv}(j)$. Similarly, a robust residual CCF is obtained by replacing $\{u_t\}$ and $\{v_t\}$ by the residual series $\{\hat{u}_t\}$ and $\{\hat{v}_t\}$. The scale parameter σ_u is estimated simultaneously from the observations using, for example, the robust scale estimator

$$\hat{\sigma}_u = \text{med}(|\hat{u}_{p+1}|, \ldots, |\hat{u}_n|)/0.6745. \tag{4}$$

Several choices for the function η are possible, such as Mallows type or Hampel type functions:

$$\eta_M(u,v) = \psi(u)\psi(v), \quad \text{Mallows}, \tag{5}$$

$$\eta_H(u,v) = \psi(uv), \quad \text{Hampel}, \tag{6}$$

where the function ψ is continuous and odd. A first example for the function ψ is the Huber family

$$\psi_H(u;c) = \text{sign}(u)\min(|u|, c); \tag{7}$$

we can interpret $\psi_H(u_t/\sigma_u; c)$ as a winsorized residual (Martin and Yohai (1985, p.138)). Another example is the bisquare family proposed by Beaton and Tukey (1974):

$$\psi_B(u;c) = \begin{cases} u(1-u^2/c^2)^2, \text{if } 0 \le |u| \le c, \\ 0, \qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{8}$$

Efficiency considerations with respect to LS estimators under a purely Gaussian model dictate the choice of the robustness constant c. See Table 1 in Bustos and Yohai (1986). Li and Hui (1994) studied the asymptotic properties of a fixed length vector of robust residual CCF.

In practical applications, there are situations where it is sufficient to robustly estimate one equation only, say the first time series, but not the second. Thus if one source is very reliable but the other one is not, we could use $\eta(u, v) = \psi(u)v$. It means that robust estimation would be used for the first series and ordinary LS estimation for the second one.

Remark. Note that robustness is not achieved uniquely with the robust CCF. If non-robust estimators (e.g., the LS estimators) are used in conjunction with (3), then if AO are present, the estimator bias can be so high that the residuals of the contaminated time series will not reflect the general dependence structure of the residuals of the uncontaminated time series. In some sense, it is similar to classical regression where a single bad leverage point can bring the LS estimators over any prespecified bound, and therefore the regression line does not reflect the general relationship between the response and the explanatory variables. See Rousseeuw and Leroy (1987, Chap. 2).

3. Test Procedure Based on All Lags

For the null hypothesis of independence, Li and Hui (1994) proposed the portmanteau test statistic

$$S^*_{RM} = \frac{n}{\hat{a}} \sum_{j=-M}^{M} \frac{n}{(n-|j|)} \gamma^2_{\hat{u}\hat{v}}(j;\eta), \tag{9}$$

a robust version of Haugh's (1976) statistics, where $\hat{a}$ is a consistent estimator of $a = E[\eta^2(u_1/\sigma_u, v_1/\sigma_v)]$. For example, when η is of the Mallows type, a consistent estimator for a is $\hat{a} = [n^{-1}\sum_{t=1}^n \psi^2(\hat{u}_t/\hat{\sigma}_u)] \times [n^{-1}\sum_{t=1}^n \psi^2(\hat{v}_t/\hat{\sigma}_v)]$. As in Haugh (1976), the correction factor $n/(n - |j|)$ in (9) leads to a better approximation by the asymptotic distribution under the null hypothesis of independence. Here we propose a robust version of Hong's statistic Q_n given by (2). It uses a kernel-based statistic. To satisfy the following assumptions.

Assumption B. The kernel $k : \mathbb{R} \rightarrow [-1, 1]$ is a symmetric function, continuous at 0, having at most a finite number of discontinuity points and such that $k(0) = 1$, $\int_{-\infty}^{\infty} k^2(z)dz < \infty$.

The test statistic is

$$Q_{Rn} = \frac{n\hat{a}^{-1} \sum_{j=1-n}^{n-1} k^2(j/m)\gamma_{\hat{u}\hat{v}}^2(j;\eta) - M_n(k)}{[2V_n(k)]^{1/2}}, \tag{10}$$

where $M_n(k)$ and $V_n(k)$ are the same quantities as in (2). The smoothing parameter m is such that $m = m(n) \rightarrow \infty$, but $m(n)/n \rightarrow 0$ as $n \rightarrow \infty$. Using the truncated uniform kernel, we retrieve a standardized version of Li and Hui's (1994) statistic. Since $\lim_{n\rightarrow\infty} m^{-1}M_n(k) = M(k) = \int_{-\infty}^{\infty} k^2(z)dz < \infty$, under some additional assumptions on k and m, $\lim_{n\rightarrow\infty} m^{-1}V_n(k) = V(k) = \int_{-\infty}^{\infty} k^4(z)dz < \infty$ and therefore $M_n(k)$ and $V_n(k)$ can be replaced by $mM(k)$ and $mV(k)$, respectively, in (10). The resulting test statistic is

$$Q_{Rn}^* = \frac{n\hat{a}^{-1} \sum_{j=1-n}^{n-1} k^2(j/m)\gamma_{\hat{u}\hat{v}}^2(j;\eta) - mM(k)}{[2mV(k)]^{1/2}}. \tag{11}$$

When $m^{-1}M_n(k) = M(k) + o(m^{-1/2})$, (11) is asymptotically equivalent in distribution to Q_{Rn}. In practice, these substitutions may lead to better finite sample approximations.

Although Q_{Rn} and Q_{Rn}^* are defined in terms of cross-correlations, there are also coherency-based statistics (in the frequency domain). For a function $g : [-\pi, \pi] \rightarrow \mathbb{C}$, the normalized L^2 norm is

$$||g||^2 = (2\pi)^{-1} \int_{-\pi}^{\pi} |g(\omega)|^2 d\omega,$$

where $|\cdot|$ denotes the modulus of a complex number. To simplify the discussion, assume that η is of the Mallows type. Then, $a^{-1/2}\gamma_{uv}(j;\eta)$ can be interpreted as the sample cross-correlation at lag j of the winsorized series $\{\psi(u_t/\sigma_u)\}$ and $\{\psi(v_t/\sigma_v)\}$. A kernel-based robust estimator of the cross-spectral density of the winsorized series is $\hat{f}_{uv}(\omega) = a^{-1/2} \sum_{j=-n+1}^{n-1} k(j/m)\gamma_{uv}(j;\eta)e^{-ij\omega}$, where $||\hat{f}_{uv}||^2 = a^{-1} \sum_{j=-n+1}^{n-1} k^2(j/m)\gamma_{uv}^2(j;\eta)$. Therefore, Q_{Rn} and Q_{Rn}^* are essentially standardized versions of the estimated coherency. See Priestley (1981) for the properties of the cross-spectral density and coherency functions.

Under the hypothesis that $\{u_t\}$ and $\{v_t\}$ are mutually independent processes, the asymptotic distribution of Q_{Rn} is the following, with convergence in law denoted $\rightarrow_L$.

Theorem 1. *Let $\{X_t\}$ and $\{Y_t\}$ be ARMA processes. Suppose that A and B are satisfied and that $m \rightarrow \infty$ and $m/n \rightarrow 0$. If the innovation processes $\{u_t\}$ and $\{v_t\}$ are independent, then $Q_{Rn} \rightarrow_L N(0, 1)$.*

The proof of Theorem 1 is written in two parts and is an adaptation of the one in Hong (1996b). First we establish the asymptotic normality of the pseudo-statistic

$$\tilde{Q}_{Rn} = \frac{na^{-1}\sum_{j=1-n}^{n-1} k^2(j/m)\gamma_{uv}^2(j;\eta) - M_n(k)}{[2V_n(k)]^{1/2}}$$

which is based on the innovation processes $\{u_t\}$ and $\{v_t\}$. To derive the asymptotic normality, we make use of Brown's (1971) Central Limit Theorem for martingale differences that can be found in many textbooks, for example Taniguchi and Kakizawa (2000, Chap. 1). The ARMA models describing the two series do not intervene in the first part since $\tilde{Q}_{Rn}$ is completely determined by the innovation series $\{u_t\}$ and $\{v_t\}$. The observed data and the estimated models are taken into account in the second part, in which it is shown that $\tilde{Q}_{Rn} - Q_{Rn} = o_p(1)$ and Theorem 1 follows. A detailed proof can be found in a technical report by Duchesne and Roy (2001).

4. Tests Based on a Subset of Lags

By making use of a general result of Li and Hui (1994) on the asymptotic distribution of a vector of robustified residual cross-correlations, we present a robust procedure for testing the hypothesis of independence based on the robustified cross-correlations at a particular lag or at a finite number of lags. We also describe a robustified version of the procedure for testing causality in mean in the Granger (1969) sense, discussed in El Himdi and Roy (1997).

4.1. Asymptotic distribution of robust cross-correlations

The theorem of Li and Hui (1994) is based on RA-estimation of the ARMA parameters and is a direct generalization of a similar result in McLeod (1979) for LS estimators. It also extends Haugh's (1976) theorem. However, when going through the Taylor series expansions involved in Haugh's proof (see also El Himdi and Roy (1997)), it is seen that Haugh's approach remains valid for any $n^{1/2}$-consistent ARMA estimators. That remark leads us to Theorem 2 below, which is slightly more general than the result of Li and Hui (1994) under the null hypothesis of independence between the two processes $\{X_t\}$ and $\{Y_t\}$.

Let $j_1, \ldots, j_m$ be a sequence of m distinct integers and consider the vector $\boldsymbol{\gamma}_{uv}^{\eta} = (\gamma_{uv}(j_1;\eta), \ldots, \ldots, \gamma_{uv}(j_m;\eta))'$.

Theorem 2. *Let $\{X_t\}$ and $\{Y_t\}$ be second-order stationary and invertible ARMA processes satisfying the assumptions of Section 2.1. Let $\{\hat{u}_t\}$ and $\{\hat{v}_t\}$ be the residual series resulting from $n^{1/2}$-consistent and robust estimation of the parameters. If the two processes are independent, then $\sqrt{n}\boldsymbol{\gamma}_{\hat{u}\hat{v}}^{\eta}$ and $\sqrt{n}\boldsymbol{\gamma}_{uv}^{\eta}$ have the same asymptotic distribution $N_m(\mathbf{0}, a\mathbf{I}_m)$, where $a = E[\eta^2(u_1/\sigma_u, v_1/\sigma_v)]$.*

A similar result is given in Haugh (1976) for two univariate time series, and in El Himdi and Roy (1997) for two multivariate time series, when $\boldsymbol{\gamma}^{\eta}_{\hat{u}\hat{v}} = \mathbf{r}_{\hat{u}\hat{v}} = (r_{\hat{u}\hat{v}}(j_1),\ldots,r_{\hat{u}\hat{v}}(j_m))'$ and the parameters are estimated by conditional LS. In the next two sections we apply Theorem 2.

4.2. Test based on the robust cross-correlation at a particular lag

Suppose that we want to test in a robust manner that the two stationary and invertible ARMA processes $\{X_t\}$ and $\{Y_t\}$ are independent. Let $\rho_{uv}(j) = \mathrm{Corr}(u_t, v_{t-j})$ be the population cross-correlation at lag j, between $\{u_t\}$ and $\{v_t\}$. Theorem 2 allows us to construct a robust version of the tests studied in Haugh (1976) and in El Himdi and Roy (1997).

Under the hypothesis of independence H_0, we have from Theorem 2 that $\sqrt{n}\gamma_{\hat{u}\hat{v}}(j_i;\eta)/\sqrt{a}$, $i = 1,\ldots,m$, are asymptotically i.i.d. $N(0,1)$. For the alternative hypothesis, $H_{1j} : \rho_{uv}(j) \neq 0$, it is natural to consider the statistic

$$S^*_R(j) = \frac{n^2}{n-|j|}\gamma^2_{\hat{u}\hat{v}}(j;\eta)/\hat{a}, \tag{12}$$

and under H_0, $S^*_R(j)$ follows a χ^2_1 distribution. The test statistic $S^*_R(j)$ is a robust version of the following test statistic studied by Haugh (1976):

$$S^*(j) = \frac{n^2}{n-|j|}r^2_{\hat{u}\hat{v}}(j). \tag{13}$$

If $n/(n-|j|)$ is replaced by one in (12) and (13), the asymptotic distribution is unchanged, but Haugh noticed from a simulation study that the exact level of $nr^2_{\hat{u}\hat{v}}(j)$ can be much smaller than the asymptotic nominal level, specially for high lags.

In practice, we are often interested in simultaneously considering several lags, for example all lags such that $|j| \leq M$, where $M \leq n-1$. Thus, the alternative hypothesis becomes $H_1^{(M)}$: There exists at least one j, $|j| \leq M$, such that $\rho_{uv}(j) \neq 0$. A global test for H_0 based on the statistics $S^*_R(j)$, $|j| \leq M$, rejects H_0 if, for at least one lag j, $S^*_R(j) > \chi^2_{1,1-\alpha_0}$. To obtain a global level α, since the $S^*_R(j)$ are asymptotically independent, the marginal level α_0 for each test must be $\alpha_0 = 1-(1-\alpha)^{1/(2M+1)}$. That method has the advantage of individual examination at each lag in a robust way. Portmanteau tests do not share that property. In Section 6, a graphical procedure as in El Himdi and Roy (1997) is presented.

4.3. Links with causality in mean

In many economic studies, it is important to identify the causality directions between two time series when they are correlated at possibly many lags. Following El Himdi and Roy (1997), the robust cross-correlation analysis described in the previous section can be pursued further for that purpose. In a general framework, the process $\{X_t\}$ does not cause $\{Y_t\}$ in the Granger (1969) sense if the minimum mean square error linear predictor of Y_t based on the information set $\{(X_s, Y_s), s \leq t-1\}$ is the same as the one based on the information set $\{Y_s, s \leq t-1\}$. In other words, $\{X_t\}$ does not cause $\{Y_t\}$ if Y_t cannot be predicted more efficiently when the information contained in $X_s, s \leq t-1$ is taken into account in addition to that in $Y_s, s \leq t-1$. More formal definitions of causality and characterizations of non-causality in vector ARMA models in terms of AR and MA parameters are given in Boudjellaba, Dufour and Roy (1992, 1994).

Pierce and Haugh (1977) obtained a characterization of the absence of causality in mean between the ARMA processes $\{X_t\}$ and $\{Y_t\}$ in terms of the cross-correlation between the two innovation processes $\{u_t\}$ and $\{v_t\}$. They showed that $\{X_t\}$ does not cause $\{Y_t\}$ if and only if $\rho_{uv}(j) = 0$, $\forall j < 0$ and $\{Y_t\}$ does not cause $\{X_t\}$ if and only if $\rho_{uv}(j) = 0$, $\forall j > 0$. Therefore, the residual CCF can be useful for detecting causality in mean directions between $\{X_t\}$ and $\{Y_t\}$.

Suppose that we want to test the null hypothesis of independence H_0 against the alternative $H_{1M}^{+} : \rho_{uv}(j) \neq 0$, for at least one j such that $1 \leq j \leq M$, where $M > 0$ is a fixed integer. If we doubt some observations, it is natural to consider robust residual cross-correlations at positive lags to define the test statistic, and if H_0 is rejected in favor of H_{1M}^{+} we will conclude that $\{Y_t\}$ causes $\{X_t\}$. As in the previous section, we can do simultaneous tests at lags $1, \ldots, M$ or employ the portmanteau type statistic

$$S_{RM}^{+*} = \sum_{j=1}^{M} S_R^{*}(j), \tag{14}$$

asymptotically χ_M^2 distribution under H_0; the null hypothesis is rejected for large values of S_{RM}^{+*}. Similarly, for the alternative hypothesis $H_{1M}^{-} : \rho_{uv}(j) \neq 0$, for at least one j such that $-M \leq j \leq -1$, the test statistic is

$$S_{RM}^{-*} = \sum_{j=-M}^{-1} S_R^{*}(j), \tag{15}$$

asymptotically χ_M^2 under H_0. If H_0 is rejected in favor of H_{1M}^{-}, we conclude that $\{X_t\}$ causes $\{Y_t\}$.

A possible interpretation of the approach leading to the test statistics (14) and (15) is as follows. Consider the Mallows case for the η function. Since we suspect outliers, we may want to perform a winsorization of the time series. Consider the following models

$$X_t^{*} = \hat{\phi}_1^{-1}(B)\hat{\theta}_1(B)u_t^{*}, \tag{16}$$

$$Y_t^{*} = \hat{\phi}_2^{-1}(B)\hat{\theta}_2(B)v_t^{*}, \tag{17}$$

where $u_t^* = \psi(u_t(\hat{\boldsymbol{\lambda}}_u)/\hat{\sigma}_u)\hat{\sigma}_u$ and $v_t^* = \psi(v_t(\hat{\boldsymbol{\lambda}}_v)/\hat{\sigma}_v)\hat{\sigma}_v$ with $(\hat{\boldsymbol{\lambda}}_u', \hat{\sigma}_u)'$ and $(\hat{\boldsymbol{\lambda}}_v', \hat{\sigma}_v)'$ denoting the robust parameter estimators of the two models as in Assumption A. Expressions (16) and (17) can be found in Bustos and Yohai (1986, Section 2.2). Using Pierce and Haugh's (1977) result, detecting causality in mean between X_t^* and Y_t^* will be in terms of the cross-correlation between the two winsorized innovation processes $\{u_t^*\}$ and $\{v_t^*\}$. The usual measure given by (13) based on the winsorized innovations u_t^* and v_t^* reduces to the robust measure defined by (12). Naturally, if no outliers are in the data, we typically have $u_t^* = u_t$, $v_t^* = v_t$, and the non-robust and robust cross-correlations will present very similar pictures as illustrated by Figure 1 of Section 6.

From a theoretical point of view, we must be cautious in doing a causality analysis based on robust residual cross-correlations. Under the hypothesis of independence between the two series, it follows from Theorem 2 that the lag j robust cross-correlation $\gamma_{\hat{u}\hat{v}}(j;\eta)$ is still a consistent estimator of $\rho_{uv}(j) = 0$. However, when $\rho_{uv}(j) \neq 0$, we know from Li and Hui (1994, Section 3) that $\gamma_{\hat{u}\hat{v}}(j;\eta)$ is in general an asymptotically biased estimator of $\rho_{uv}(j)$. Therefore, it may well happen, due to the bias, that $\gamma_{\hat{u}\hat{v}}(j;\eta)$ be close to zero even if $\rho_{uv}(j) \neq 0$. In such a situation, no causality direction is detected with $\gamma_{\hat{u}\hat{v}}(j;\eta)$. In the context of causality analysis, the behavior of the robust auto- or cross-correlations, and in particular their asymptotic bias, when the corresponding theoretical correlations are different from zero must be studied further but we do not do so here. In practical applications, it is expected that the bias will be small since without outliers very few residuals, if any, will be downweighted in (3). Typically, the usual and robustified cross-correlations will be very similar with outlier-free samples. Furthermore, when outliers are observed, the robustified measures should be considerably less biased than the usual cross-correlations based on the LS estimators of the ARMA parameters. A more formal approach, which could probably avoid the bias induced by the robustified cross-correlation measures in causality in mean analysis, could rely on the general concept of causality discussed in Granger (1980).

The procedure described here for causality in mean analysis does not constitute a statistical test in the usual sense, since the hypothesis of non-causality which is the hypothesis of interest stands for the alternative hypothesis rather than the null. It is however a robustified version of a similar approach used by, among others, Pierce (1977) and Pierce and Haugh (1977) to identity causality in mean directions between univariate time series. From a statistical point of view, we would like to directly consider the null hypothesis $H_0^+ : \rho_{uv}(j) \neq 0, j > 0$, or $H_0^- : \rho_{uv}(j) \neq 0, j < 0$ versus the simple negation of H_0^+ or H_0^-. Li and Hui (1994) give the asymptotic distribution of the vector of residual robust cross-correlations $\hat{\boldsymbol{\gamma}}_{\hat{u}\hat{v}}^{\eta}$ in the general case of two correlated univariate ARMA series.

The asymptotic covariance matrix is quite involved since it depends on the true models describing the two series and except for the special case of only instantaneous causality between the two processes $\{X_t\}$ and $\{Y_t\}$, it seems difficult to exploit in practice.

5. Simulation Results

From a practical point of view, it is natural to inquire after the finite sample properties of the proposed test statistics, in particular their exact level and power whether or not there is outliers in the series under study. To partially answer that question, a small Monte Carlo simulation was conducted. In addition to the test statistics discussed in the preceding sections, the following modified versions of Haugh's (1976) and Hong's (1996a) statistics S_M^* and Q_n^* were also included in the experiment:

$$S_M^* = n \sum_{j=-M}^{M} \frac{n}{(n-|j|)} r_{\hat{u}\hat{v}}^2(j), \tag{18}$$

$$Q_n^* = \frac{n \sum_{j=1-n}^{n-1} k^2(j/m) r_{\hat{u}\hat{v}}^2(j) - mM(k)}{[2mV(k)]^{1/2}}. \tag{19}$$

In our simulation experiment, the empirical levels of S_M^* and Q_n^* were closer to the nominal levels than those of S_M and Q_n, respectively. However, the empirical powers of S_M^* and Q_n^* were similar to those of S_M and Q_n, respectively. For these reasons, we only report the results for (18) and (19).

5.1. Description of the experiment

To compare the omnibus robust statistic Q_{Rn}^* to its non-robust counterpart Q_n^* and also to the portmanteau statistics S_M^* and S_{RM}^*, independent realizations were generated from the following bivariate Gaussian VARMA(1,1) process:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix} - \begin{pmatrix} 0.0 & \theta \\ \theta & 0.0 \end{pmatrix} \begin{pmatrix} u_{t-1} \\ v_{t-1} \end{pmatrix}, t \in \mathbb{Z}. \tag{20}$$

The covariance matrix of $(u_t, v_t)'$ is

$$\Sigma = \begin{pmatrix} 1.0 & \rho \\ \rho & 1.0 \end{pmatrix}. \tag{21}$$

We considered (i) $(\theta, \rho) = (0.0, 0.0)$, (ii) $(\theta, \rho) = (0.0, 0.2)$ and (iii) $(\theta, \rho) = (0.25, 0.0)$. The first case corresponds to the situation where two independent AR(1) series are generated, and that situation allows us to study the level of the tests. Under (ii), there is instantaneous correlation between the two innovation series of the two marginal processes. The case (iii) is similar to the second experiment of Li and Hui (1994, p.108), where the two innovation series of the marginal processes are correlated at lags ± 1. The cases (ii) and (iii) allow us to compare the power of the various tests.

For each kernel, three different rates m were employed: $\lfloor \log(n) \rfloor$, $\lfloor 3.5n^{0.2} \rfloor$ and $\lfloor 3n^{0.3} \rfloor$ ($\lfloor a \rfloor$ denotes the integer part of a). These rates are discussed in Hong (1996c, p.849). They lead respectively to the values $m = 5, 8, 12$ for the series length $n = 100$, and to $m = 5, 9, 15$ with $n = 200$. We used the same rates for the portmanteau statistics S_M^* and S_{RM}^*, that is $M = m$. Kernels used in the calculation of the omnibus statistics Q_n^* and Q_{Rn}^* are given in Table 1. For more details on the different kernels, see Priestley (1981).

Table 1. Kernels used in the empirical study.

Truncated uniform kernel (TR):	$k(z) = \begin{cases} 1 \text{ if } \lvert u \rvert \leq 1, \\ 0 \text{ elsewhere;} \end{cases}$
Bartlett (BAR):	$k(z) = \begin{cases} 1 - \lvert z \rvert & \text{if } \lvert z \rvert \leq 1, \\ 0 & \text{elsewhere;} \end{cases}$
Daniell (DAN):	$k(z) = \dfrac{\sin(\pi z)}{\pi z}, \ z \in \mathbb{R}.$

To investigate the effect of outliers on the robust and non-robust statistics, 10,000 realizations were generated from the Gaussian VAR(1) model (20) for each series length ($n = 100, 200$). For each realization, the following three scenarios were considered.

Scenario 1: No contamination.
Scenario 2: For $n = 100$, we subtracted 10 from observations X_{26} and Y_{76}. Furthermore, for $n = 200$, we added 10 to X_{101} and Y_{151} and subtracted 10 from X_{176}.
Scenario 3: In addition to scenario 2, for $n = 100$, we added 10 to X_{51} and Y_{51} and, for $n = 200$, we subtracted 10 from X_{126} and Y_{126}.

Similar scenarios were employed in Li (1988) and Li and Hui (1994). The first allows us to evaluate the performance of the statistics in the context of Gaussian VAR(1) and VARMA(1,1) series. The second permits us to evaluate the performance of the procedures when outliers occur at different points in time. Finally, in the third scenario, outliers occur in both series at the same points in time.

For the robust statistics, we chose the Mallows type functions η defined by (5) and two ψ functions were employed: Huber (7) and bisquare (8) families.

The tuning constant c in the Huber family is 1.65 and in the bisquare family is 5.58. These robustness constants give an efficiency of 95% with respect to the LS estimators under a perfectly observed Gaussian time series.

In the level study, 10,000 independent realizations were generated from (20) and (21) with $(\theta, \rho) = (0.0, 0.0)$ for each value of n. Computations were made in the following way, with computer programs written in Fortran 77.

(1) The Gaussian white noise $(u_t, v_t)'$, $t = 1, \ldots, n$, was generated using the subroutine G05EZF from the NAG library.
(2) Initial values $(X_0, Y_0)'$ were generated from the exact bivariate Gaussian distribution using Ansley's (1980) algorithm. The values $(X_t, Y_t)', t = 1, \ldots, n$, were obtained by solving the difference equation (20).
(3) For each realization, univariate AR(1) models were separately estimated for each of the two series and the residual series $\{\hat{u}_t, t = 1, \ldots, n\}$ and $\{\hat{v}_t, t = 1, \ldots, n\}$ were obtained. The autoregressive parameter was estimated by LS for the non-robust statistics and the RA estimator was obtained from the iterative algorithm described in Bustos and Yohai (1986, Section 2) for the robust statistics. The Mallows type function η combined with Huber and bisquare functions ψ were used. The three scenarios for outliers were applied to each realization, which leads to nine distinct estimates and therefore to nine residual series $\{(\hat{u}_t, \hat{v}_t)', t = 1, \ldots, n\}$.
(4) For each realization, the test statistics S_M^* and Q_n^* were computed from the LS residuals. The value of S_{RM}^* and Q_{Rn}^* were obtained from the RA residuals.
(5) For each series of length n and for nominal level 5%, we obtained for each statistic the number of rejections of the null hypothesis of non-correlation between the two series, based on 10,000 realizations. The standard error of the number of rejections is 21.8 for 5%. For example, at the nominal level 5%, the observed number of rejections is not significantly different from the expected number of rejections if it lies in the interval [458, 542] at the significance level $\alpha = 0.05$ and in the interval [444, 556] at the significance level $\alpha = 0.01$.

The power study was conducted in a similar way except that in case (ii) the correlation between u_t and v_t is $\rho = 0.2$ and in case (iii) a VARMA(1,1) is generated. Note that it is appropriate to estimate AR(1) models under the null hypothesis and under the two alternatives considered.

For the kernel-based test statistics, we give the results for the Bartlett, Daniell and truncated uniform kernels (see Table 1) for Q_n^* and Q_{Rn}^*. The Daniell kernel is optimal (greatest asymptotic power) kernel in a certain class of smooth kernels (Hong (1996a), Section 4). Other complementary simulation results for other statistics, kernels and significance levels can be found in Duchesne (2000, Chap. 4). We give here the more representative results.

5.2. Discussion of the level study

Since the nominal level is 5%, all the numbers in Table 2 must be compared to 500. In Tables 2, 3 and 4, $S^*_{RM}(H)$ and $S^*_{RM}(b)$ represent the statistic S^*_{RM} computed with the Huber and bisquare ψ functions respectively. Under scenario 1 (no outlier), the levels of the portmanteau statistics S^*_M and S^*_{RM} are very well controlled. Hong's statistic Q^*_n tends to overreject. When $m = 5$, its empirical level varies around 7% with the BAR and DAN kernels, around 8% with the TR kernel and becomes closer to 5% as m increases. A similar trend was observed by Hong (1996a). The behavior of the robustified version Q^*_{Rn} is similar to that of Q^*_n, irrespective of the ψ function.

Table 2. Number of rejections under the null hypothesis based on 10,000 realizations at the 5% nominal level. For each value of $m = M$, the three lines correspond to the three scenarios.

				Q^*_n			Q^*_{Rn}					
							Huber			bisquare		
m	S^*_M	$S^*_{RM}(H)$	$S^*_{RM}(b)$	BAR	DAN	TR	BAR	DAN	TR	BAR	DAN	TR
$n=100$												
5	443	454	457	753	657	835	776	657	842	769	664	828
	49	471	492	175	119	121	737	625	828	748	649	850
	3159	552	472	9480	8632	4754	1101	894	988	740	651	822
8	469	460	468	673	555	645	670	542	651	674	544	647
	29	464	485	94	70	42	650	552	636	670	556	658
	1215	539	454	8581	6722	1876	960	715	750	661	545	654
12	488	487	496	575	432	473	566	425	470	580	438	477
	19	510	471	57	31	17	586	427	500	590	437	461
	309	544	484	6902	4073	407	767	539	535	565	436	474
$n=200$												
5	504	504	490	781	669	899	764	666	898	779	687	901
	27	473	484	156	95	73	714	627	865	762	677	900
	8188	744	493	9997	9977	9101	1673	1365	1307	786	678	933
9	494	491	482	678	590	737	677	585	709	676	582	732
	12	497	487	62	45	28	636	564	696	671	575	715
	4269	658	487	9936	9740	5581	1348	1088	927	701	609	717
15	485	475	465	614	513	546	602	501	525	607	489	531
	7	506	504	28	31	10	578	482	557	589	491	554
	1069	611	495	9498	9105	1437	1088	827	687	595	478	556

Under scenarios 2 and 3, it is immediately seen that the unrobustified statistics S_M^* and Q_n^* are very sensitive to outliers. In general, they dramatically underreject under scenario 2 and overreject under scenario 3. In contrast, the empirical level of the robustified statistics S_{RM}^* under scenarios 2 and 3, those of Q_{Rn}^* with either Huber or the bisquare function ψ under scenario 2, and those of Q_{Rn}^* with the bisquare function are similar to the ones observed under scenario 1 and are therefore reasonably close to 5 %. However, Q_{Rn}^* with Huber function overrejects considerably under scenario 3, specially for small values of m, with the three kernels and for all values of m with the kernel BAR. For that reason, the bisquare function seems preferable in practice.

Table 3. Number of rejections under the alternative $(\theta, \rho) = (0.0, 0.2)$ in (20) and (21), based on 10,000 realizations at the 5 % nominal level and using the empirical critical values for scenario 1. For each value of $m = M$, the three lines correspond to the three scenarios.

				Q_n^*			Q_{Rn}^*					
							Huber			bisquare		
m	S_M^*	$S_{RM}^*(H)$	$S_{RM}^*(b)$	BAR	DAN	TR	BAR	DAN	TR	BAR	DAN	TR
$n = 100$												
5	2076	1939	1914	4033	3400	2153	3890	3242	2001	3861	3229	1968
	136	1666	1839	621	404	143	3241	2700	1709	3565	2995	1901
	6568	2626	1856	9894	9692	6864	5226	4526	2727	3702	3131	1919
8	1591	1586	1549	3520	2871	1668	3421	2746	1642	3402	2733	1621
	79	1394	1532	403	243	79	2814	2253	1450	3156	2505	1602
	3495	2107	1541	9681	9053	3921	4681	3823	2194	3296	2588	1617
12	1382	1315	1310	3067	2488	1460	2918	2325	1417	2926	2311	1398
	42	1195	1250	247	127	44	2417	1928	1264	2706	2186	1320
	1279	1750	1278	9141	7582	1712	4045	3231	1875	2802	2223	1374
$n = 200$												
5	4039	3881	3881	7145	6394	4103	6919	6157	3945	6902	6137	3933
	165	3029	3528	988	699	170	5886	5145	3075	6473	5666	3584
	9824	5266	3621	10000	10000	9837	8405	7785	5341	6627	5864	3676
9	3132	2954	3003	6292	5376	3218	6074	5156	3055	6055	5147	3083
	82	2361	2752	549	346	83	5016	4114	2440	5625	4676	2832
	8389	4098	2848	9999	9994	8544	7660	6727	4224	5768	4859	2908
15	2429	2325	2337	5429	4377	2522	5169	4192	2428	5178	4202	2422
	23	1889	2125	260	235	25	4113	3362	1965	4732	3836	2217
	4237	3173	2234	9974	9937	4765	6692	5747	3314	4912	3964	2317

Globally as a function of m, the sizes of the robust statistics get better as m increases. With respect to the series length, there is no obvious improvement when n goes from 100 to 200.

Table 4. Number of rejections under the alternative $(\theta, \rho) = (0.25, 0.0)$ in (20) and (21), based on 10,000 realizations at the 5 % nominal level and using the empirical critical values for scenario 1. For each value of $m = M$, the three lines correspond to the three scenarios.

				Q_n^*			Q_{Rn}^* Huber			Q_{Rn}^* bisquare		
m	S_M^*	$S_{RM}^*(H)$	$S_{RM}^*(b)$	BAR	DAN	TR	BAR	DAN	TR	BAR	DAN	TR
$n=100$												
5	5700	5378	5328	7082	7530	5815	6801	7275	5471	6752	7260	5440
	620	4544	4897	1461	1635	654	5928	6445	4626	6254	6775	5006
	4123	5459	4569	9215	8758	4429	6806	7319	5557	5964	6503	4658
8	4523	4402	4315	7141	7058	4711	6867	6808	4554	6846	6765	4515
	306	3738	3938	1316	1194	327	5965	5882	3864	6369	6282	4117
	1628	4465	3674	8501	7365	1876	6886	6839	4610	6068	5991	3837
12	3678	3509	3472	6781	6417	3967	6490	6106	3730	6490	6126	3730
	164	2940	3133	986	770	183	5583	5257	3141	6002	5661	3372
	506	3606	2931	7185	5188	669	6513	6171	3826	5709	5348	3146
$n=200$												
5	9204	9037	9038	9729	9803	9224	9647	9737	9065	9654	9726	9055
	1589	8261	8628	3571	3796	1623	9213	9369	8298	9426	9531	8658
	9018	9039	8398	9990	9981	9092	9648	9736	9070	9292	9434	8425
9	8422	8169	8185	9689	9654	8487	9614	9558	8239	9607	9561	8271
	671	7102	7595	2964	2795	706	9145	9056	7205	9369	9296	7675
	5630	8095	7266	9952	9871	5897	9594	9547	8204	9211	9141	7359
15	7335	6966	6998	9525	9304	7487	9398	9175	7152	9412	9172	7164
	258	5883	6325	2037	1935	275	8785	8441	6080	9090	8797	6499
	1864	7019	6024	9689	9598	2138	9394	9204	7192	8914	8595	6200

5.3. Discussion of the power study

The number of rejections under the alternative hypotheses $(\theta, \rho) = (0.0, 0.2)$, $(\theta, \rho) = (0.25, 0.0)$ in (20) and (21) based on 10,000 realizations at the 5 % nominal level are reported in Table 3 and 4. We used the empirical critical values obtained from the level study under the first scenario. The powers of S_M^* and Q_n^*

under scenarios 2 and 3 are of limited interest since these statistics are severely affected by outliers. The only scenario for which all the tests are comparable is the first. In particular, there we compare the powers of robust and non-robust tests when there are no outliers. In addition, a satisfactory test should have an empirical power under scenarios 2 and 3 similar to the one obtained under the first one.

It is interesting to note that the price to pay for using a robustified statistic rather than its unrobustified version is rather small, at least for the two alternatives considered in this study. Indeed, as expected, the power of Q^*_{Rn} is smaller than that of Q^*_n, but the difference is rather small and the same can be said of S^*_{RM} in comparison with S^*_M. Since the behavior of Q^*_n was already discussed in Hong (1996a), we focus on its robustified version Q^*_{Rn}.

Among the three kernels, the BAR kernel is in general more powerful than the DAN kernel under the two alternatives, with the exception of the second alternative with $m = 5$ where the DAN kernel is more powerful. This observation does not contradict the optimality result of Hong (1996a) which says that the DAN kernel is optimal in a certain class of smooth kernels, since the BAR kernel is not in that class. The DAN and BAR kernels are considerably more powerful than the TR kernel, which is essentially the robust test S^*_{RM}. Note that the empirical power of Q^*_{Rn} with the TR kernel differs slightly from the power of the robust test of Li and Hui (1994). Very likely, the correction factor $n/(n - |j|)$ is one of the reasons. This illustrates that more powerful tests than the robust test of Li and Hui (1994) can be obtained by adopting a kernel different from the TR kernel.

The two ψ functions lead to similar powers in scenarios 1 and 2, specially for large m. In general under the two alternatives, it seems that the bisquare function gives the best results. We see from Table 3 that the Huber function leads to larger powers than the bisquare function. It is very likely related to the fact that the empirical levels with contaminated series are much larger than those with non-contaminated series.

As a function of m, the power of Q^*_{Rn} decreases considerably as m increases except with the Huber function in scenario 3. As for Hong's statistic Q^*_n, the power of Q^*_{Rn} increases very slightly when n goes from 100 to 200. Finally, among the three kernels, the TR kernel was the least powerful but still more powerful than Li and Hui's (1994) robust portmanteau statistics S^*_M. For $n = 100$, using the BAR kernel and a given ψ function, the robust test Q^*_{Rn} is at least twice as efficient than S^*_{Rn} for a given m under case (ii). Under case (iii), Q^*_{Rn} has a greater power than S^*_{Rn} by 27% when $m = 5$, and the power becomes greater by 87% when $m = 12$. For $n = 200$, the differences are smaller but the tests Q^*_{Rn} still dominated S^*_{Rn}, especially for large m.

In summary, we have seen that the levels of Q_n^* are sensitive to additive outliers, while those of Q_{Rn}^* are reasonably well controlled. In our empirical investigation, the loss of power in using Q_{Rn}^* rather than Q_n^* is not too important under a perfectly observed ARMA process. The empirical power of Q_{Rn}^* is higher than that of S_{Rn}^*, using a kernel different from the truncated uniform kernel. The empirical powers of the robust statistics Q_{Rn}^* are reasonably close under the three scenarios. If the outliers are at the same point in time (scenario 3), the bisquare function seems preferable to the Huber function.

6. Application

6.1. The data

We illustrate here the new robust tests for non-correlation with a data set coming from the financial literature. It was analyzed by Brockwell and Davis (1996, Chap. 7) who aimed at developing a stationary bivariate time series model. However, in practice, before deciding to build a multivariate model, a first important step is to check for cross-correlation between the series. If we conclude that they are indeed dependent, it is then appropriate to develop a multivariate model.

The variables are the closing values of the Dow-Jones Index of stocks (D_t) on the New York Stock Exchange and the closing values of the Australian All-ordinaries Index (A_t) of Share Prices during 251 successive trading days from September 13, 1993 to August 26th, 1994. See Brockwell and Davis (1996, p.219) for a picture of the two series. For the purpose of this study, we assume that the time series $\{D_t\}$ is obtained from a reliable source but that there are possibly some outliers in the time series $\{A_t\}$. Therefore, we robustly estimate only the second time series. We admit that this assumption is somewhat arbitrary, but it will be useful for our illustration and discussion.

According to the efficient market hypothesis, these two series should behave individually as random walks with uncorrelated increments. In order to have stationary series, the data are transformed as percentage relative price changes, that is $(X_t, Y_t)' = 100((D_t - D_{t-1})/D_{t-1}, 100(A_t - A_{t-1})/A_{t-1})'$, for $t = 1, \ldots, 250$. The sample autocorrelations of each series indicate that both of them can be considered as white noise since all their autocorrelations are within, or very close to, the significant limits $\pm 1.96 n^{-1/2}$ (Brockwell and Davis (1996), Example 7.1.1). The sample cross-correlations are also insignificant except at lag -1 where $r_{XY}(-1) = 0.46$, which indicates that X_{t-1} and Y_t are dependent.

6.2. Correlation analysis

Since stock changes price may display strong volatility clustering, we performed tests for ARCH effects. We used the one-sided test of Hong (1997), based on a kernel-based spectral density estimator at the zero frequency. We used the Bartlett kernel with a smoothing parameter chosen by cross-validation. We postulated $X_t = h_{1t}^{1/2}\xi_{1t}$ and $Y_t = h_{2t}^{1/2}\xi_{2t}$, where h_{it}, $i = 1, 2$ denote the conditional variances and $\{\xi_{1t}\}$ and $\{\xi_{2t}\}$ are i.i.d. processes. We found strong evidence of ARCH effects in the time series $\{X_t\}$. No ARCH effects were found in $\{Y_t\}$.

The test procedures described in the previous sections can be directly applied to the two series $\{X_t/h_{1t}^{1/2}\}$ and $\{Y_t/\sigma_Y\}$ since there is no significant serial autocorrelation. We formulated a GARCH(1,1) process for the conditional variance h_{1t} using the S-PLUS GARCH software. The values of the statistics $S^*(j)$ defined by (13), $|j| \leq 12$, are presented in Figure 1 (left side). With tests at individual lags, we strongly reject the hypothesis of non-correlation at lag $j = -1$. We also reject with simultaneous tests at lags $-6, \ldots, +6$ and even at lags $-12, \ldots, +12$ at the global level $\alpha = 0.05$. The values of the robust statistics $S_R^*(j)$, $|j| \leq 12$, are shown in the right side of Figure 1. Again, the non-correlation hypothesis is rejected by the test at the individual lag $j = -1$ and by the simultaneous tests at the global level $\alpha = 0.05$.

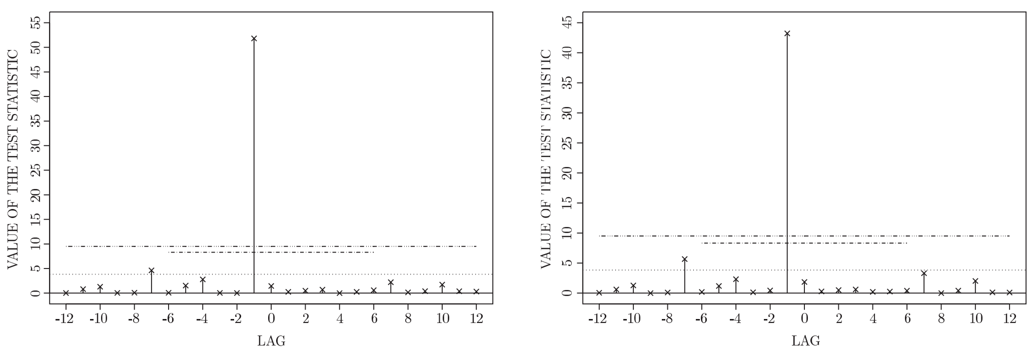

FIgure 1. Values of the statistics $S^*(j)$ (left side) and of $S_R^*(j)$ (right side) for different lags j. The horizontal dotted lines represent the marginal critical value at the level $\alpha = 0.05$. The dashed lines give the critical values at the global level $\alpha = 0.05$, for simultaneous tests at lags $j = -6, \ldots, 6$ and $j = -12, \ldots, 12$.

In order to take into account all possible lags, Hong's test Q_n^* and its robustified version Q_{Rn}^* where carried out. The statistic Q_{Rn}^* was calculated with $\eta(u, v) = u\psi_B(v)$ (the tuning constant was $c = 5.58$ as in Section 5), the Daniell kernel and the values $5, 9, 16$ for m that correspond to the three rates employed in Section 5. The scale parameter σ_Y was estimated with the median estimator defined by (4). The values of the statistics are reported in Table 4. At the 1% significance level, these values must be compared to the 99th quantile of the $N(0, 1)$ distribution and, again, the hypothesis of non-correlation is rejected.

Since there are no apparent outliers in the two time series under study, we have replaced Y_{16} by -9.0 and Y_{106} by $+9.0$ which gives rise to two additive outliers in the Australian series. The contaminated series $\{X_t\}$ and the series $\{Y_t\}$ are shown in Figure 2. Visually, the behavior of the Australian series is suspect, although comparable to many financial time series encountered in practice. The new series $\{Y_t\}$ can still be considered as a white noise and the cross-correlation at lag $j = -1$ is now 0.12 rather that 0.46. We did tests for ARCH effects in the contaminated time series and such effects were still not present. For these new series, we have redone the correlation analysis. With the non-robust procedure, no significant correlation is detected by the simultaneous tests at lags $j = -12, \dots, 12$ and $j = -6, \dots, 6$, or with Hong's statistic whose values are reported in Table 5. In fact, no cross-correlation is detected by the tests at individual lags, Figure 3 (left side). With the robust procedure illustrated in Figure 3 (right side) for the statistic $S_R^*(j)$ and in Table 5 for the statistics Q_{Rn}^*, the hypothesis of non-correlation is strongly rejected once again.

Table 5. Values of the statistics Q_n^* and Q_{Rn}^* for the data and for the data contaminated by two outliers.

	Real data		Contaminated data	
m	Q_n^*	Q_{Rn}^*	Q_n^*	Q_{Rn}^*
5	16.65	14.03	0.31	11.63
9	13.42	11.22	-0.37	9.22
16	10.26	8.69	-1.05	7.05

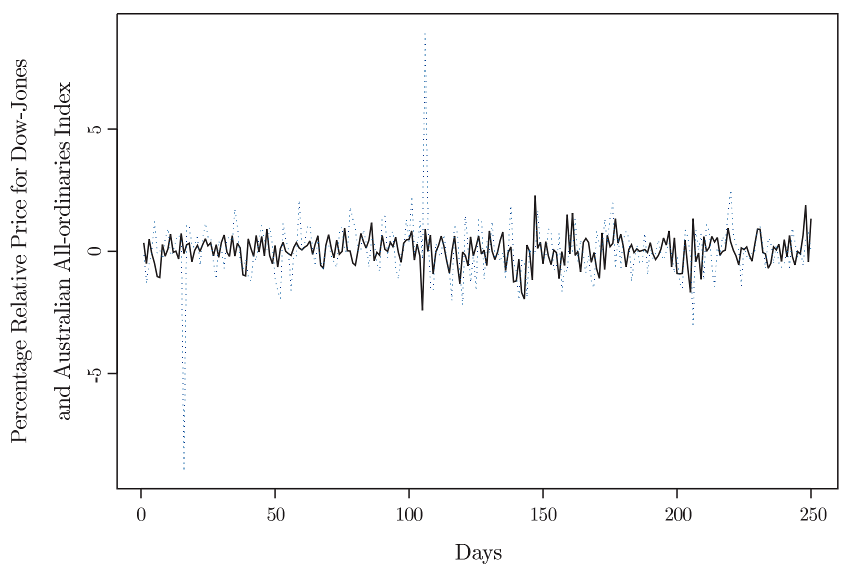

Figure 2. The two variables X_t and Y_t with two outliers. The dotted line represents the Australian series.

6.3. A descriptive causality in mean analysis

Once the hypothesis H_0 of non-correlation between the two series is rejected, it is natural to inquire into the directions of causality between the two variables, especially in economic studies. In Section 4.3, we have seen that $\{X_t\}$ does not cause $\{Y_t\}$ if and only if $\rho_{uv}(j) = 0$, $\forall j < 0$, and $\{Y_t\}$ does not cause $\{X_t\}$ if and only if $\rho_{uv}(j) = 0$, $\forall j > 0$, where $\{u_t\}$ and $\{v_t\}$ are the two corresponding innovation processes. In the first part of the correlation analysis of the previous section, we have tested H_0 with statistics based on individual lags and it is immediately seen from Figure 1 that $\rho_{uv}(-1)$ is significantly different from zero and all the other lags, the values of the test statistics are either smaller than the 5% critical value or very close of it. Figure 1 leads us to speculate that the Dow-Jones Index influenced the Australian Index and that the Australian Index did not influence the Dow-Jones. At least two factors may explain that conclusion: the relative size of the two economies and the timing (open and closing hours) of the two stock markets.

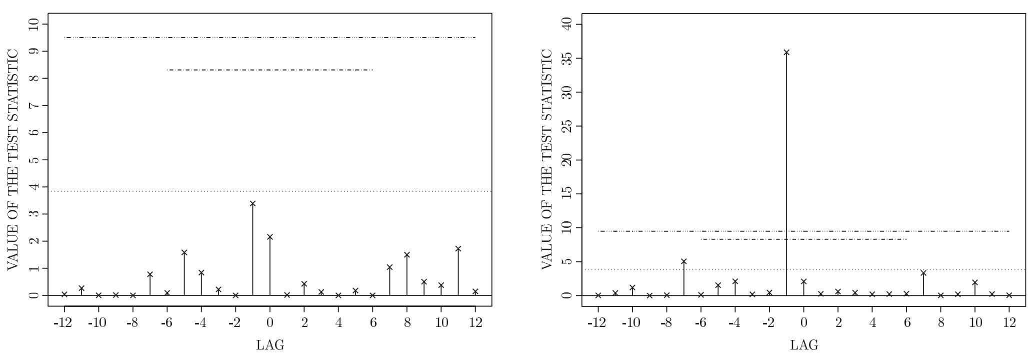

Figure 3. Values of the statistics $S^*(j)$ (left side) and of $S^*_R(j)$ (right side) for different lags j with the contaminated data. The horizontal dotted line represents the marginal critical value at the level $\alpha = 0.05$. The dash lines give the critical values at the global level $\alpha = 0.05$, for simultaneous tests at lags $j = -6, \ldots, 6$ and $j = -12, \ldots, 12$.

In order to simultaneously consider many lags, we can employ the statistic $S^{+*}_M = \sum_{j=1}^{M} S^*(j)$ or $S^{-*}_M = \sum_{j=-M}^{-1} S^*(j)$, asymptotically χ^2_M under H_0. For H_0 against H^+_{1M}, the p-values of S^{+*}_M for various values of M are presented in Table 6. For H_0 against H^-_{1M}, similar results are given for S^{-*}_M. Thus, for all the values of M considered, H_0 is rejected against H^-_{1M} and not rejected against H^+_{1M}. These tests confirm the conclusion deduced from the graphical analysis. The robust statistics S^{+*}_{RM} and S^{-*}_{RM} defined by (14) and (15), respectively, were also employed for these data with the same conclusion.

The causality analysis was redone with the contaminated data and the resulting p-values are shown in Table 6. With the non-robust statistics, H_0 is not rejected against H_{1M}^{+} as before. However, against H_{1M}^{-}, H_0 is not rejected at the 5% level for any values of M. This contrasts strongly with the robust statistics since, with them, we retrieve the conclusions obtained with the true data.

Table 6. p-values of the non-robust tests S_M^{+*}, S_M^{-*} and of the robust tests S_{RM}^{+*} and S_{RM}^{-*} for the real data and for the contaminated data.

	Real data				Contaminated data			
M	S_M^{+}	S_M^{-}	S_{RM}^{+}	S_{RM}^{-}	S_M^{+}	S_M^{-}	S_{RM}^{+}	S_{RM}^{-}
1	0.61	0^+	0.60	0^+	0.89	0.07	0.61	0^+
2	0.69	0^+	0.68	0^+	0.80	0.18	0.65	0^+
3	0.69	0^+	0.71	0^+	0.90	0.31	0.73	0^+
6	0.89	0^+	0.89	0^+	0.99	0.41	0.92	0^+
12	0.82	0^+	0.77	0^+	0.91	0.84	0.80	0^+

Note: 0^+ denotes a positive number smaller than 10^{-5}.

Acknowledgements

The authors would like to thank a co-editor and the two referees for helpful comments. We are most grateful to one referee who went well beyond the call of duty in providing constructive remarks and suggestions, leading to a significant improvement of the paper. We acknowledge helpful discussions with Simon Lalancette from the Finance Department at HEC Montreal with respect to Section 6. This work was supported by grants from the Natural Science and Engineering Research Council of Canada and the Institut de Finance Mathématique de Montréal IFM^2. This work was supported by grants from the Natural Science and Engineering Research Council of Canada, the Fonds FCAR (Government of Québec) and the Network of Centres of Excellence on the Mathematics of Information Technology and Complex Systems.

References

Akaike, H. and Kitagawa, G. (1999). *The Practice of Time Series Analysis*. Springer, Berlin.

Ansley, C. F. (1980). Computation of the theoretical autocovariance function for a vector ARMA process. *J. Statist. Comput. Simulation* **12**, 15-24.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16**, 147-185.

Ben, M. G., Martinez, E. J. and Yohai, V. J. (1999). Robust estimation in vector autoregressive moving-average models. *J. Time Ser. Anal.* **20**, 381-399.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1994). Simplified conditions for noncausality between vectors in multivariate ARMA models. *J. Econometrics* **63**, 271-287.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992). Testing causality between two vectors in multivariate autoregressive moving average models. *J. Amer. Statist. Assoc.* **87**, 1082-1090.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis. Forecasting and Control*. 3rd edition. Prentice Hall, Englewood Cliffs, NJ.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.* **65**, 1509-1526.

Brockwell, P. J. and Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. Springer, Berlin.

Brown, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42**, 59-66.

Bustos, O. H. and Yohai, V. J. (1986). Robust estimates for ARMA models. *J. Amer. Statist. Assoc.* **81**, 155-168.

Bustos, O., Fraiman, R. and Yohai, V. J. (1984). Asymptotic Behaviour of the Estimates Based on Residual Autocovariances for ARMA Models. In *Robust and Nonlinear Time Series Analysis* (Edited by J. Franke, W. Haerdle and D. Martin), 26-49, Springer, Berlin.

Carroll, R. J. and Welsh, A. H. (1988). A note on asymmetry and robustness in linear regression. *Amer. Statist.* **42**, 285-287.

Derby, L. and Martin, R. D. (1979). Robust estimation of the first-order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**, 140-146.

Duchesne, P. (2000). Quelques contributions en théorie de l'échantillonnage et dans l'analyse des séries chronologiques multidimensionnelles. Unpublished Ph.D. thesis, Department of Mathematics and Statistics, Université de Montréal.

Duchesne, P. and Roy, R. (2001). Robust tests for independence of two time series. Technical Report CRM-2751, Université de Montréal.

El Bantli, F. and Hallin, M. (2001). Asymptotic behaviour of M-estimators in AR(p) models under nonstandard conditions. *Canad. J. Statist.* **29**, 155-168.

El Himdi, K. and Roy, R. (1997). Tests for noncorrelation of two multivariate ARMA time series. *Canad. J. Statist.* **25**, 233-256.

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

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424-438.

Granger, C. W. J. (1980). Testing for causality: a personal viewpoint. *J. Economic Dynamics and Control* **2**, 329-352.

Hallin, M., Jurevcková, J., Picek, J. and Zahaf, T. (1999). Nonparametric tests of independence of two autoregressive time series based on autoregression rank scores. *J. Statist. Plann. Inference* **75**, 319-330.

Hallin, M. and Saidi, A. (2001). Testing independence and causality between multivariate ARMA time series. Working paper, ISRO, Universite Libre de Bruxelles.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *J. Amer. Statist. Assoc.* **71**, 378-385.

Hong, Y. (1996a). Testing for independence between two covariance stationary time series. *Biometrika* **83**, 615-625.

Hong, Y. (1996*b*). A separate mathematical appendix for 'Testing for independence between two covariance stationary time series'. Mimeo, Department of Economics and Department of Statistical Sciences, Cornell University.

Hong, Y. (1996*c*). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837-864.

Hong, Y. (1997). One-sided testing for conditional heteroscedasticity in time series models. *J. Time Ser. Anal.* **18**, 253-277.

Hosking, J. (1980). The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.* **75**, 602-608.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*. Second edition. Wiley, New York.

Koch, P. D. and Yang, S. S. (1986). A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. *J. Amer. Statist. Assoc.* **81**, 533-544.

Li, W. K. (1988). A goodness-of-fit test in robust time series modelling. *Biometrika* **75**, 355–361.

Li, W. K. and Hui, Y. V. (1994). Robust residual cross correlation tests for lagged relations in time series. *J. Statist. Comput. Simulation* **49**, 103–109.

Martin, R. D. and Yohai, V. J. (1985). Robustness in time series and estimating ARMA models. In *Handbook of Statistics,* 5, *Time Series in the Time Domain* (Edited by E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 119-155, North-Holland, Amsterdam.

McLeod, A. I. (1979). Distribution of the residual cross-correlation in univariate ARMA time series models. *J. Amer. Statist. Assoc.* **74**, 849-855.

Pham, D. T., Roy, R. and Cédras, L. (2003). Tests for non-correlation of two cointegrated ARMA time series. *J. Time Ser. Anal.*, in press.

Pierce, D. A. (1977). Relationships – and the lack thereof – between economic time series, with special reference to money and interest rates. *J. Amer. Statist. Assoc.* **72**, 11-22. (C/R: p22-26).

Pierce, D. A. and Haugh, L. D. (1977). Causality in temporal systems: characterizations and survey. *J. Econometrics* **5**, 265-293.

Priestley, M. B. (1981). *Spectral Analysis and Time Series: Univariate Series*, Vol. 1. Academic Press, London.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

Taniguchi, M. and Kakizawa, Y. (2000). Asymptotic Theory of Statistical Inference for Time Series. Springer, Berlin.

Université de Montréal, Departement de mathematiques et de statistique, CP 6128 Succ Centre-Ville, Montreal QC H3C 3J7, Canada.

E-mail: duchesne@dms.umontreal.ca

Université de Montréal, Departement de mathematiques et de statistique, CP 6128 Succ Centre-Ville, Montreal QC H3C 3J7, Canada.

E-mail: royr@dms.umontreal.ca

(Received August 2001; accepted April 2003)