Appl. Statist. (1992)
41, *No.* 1, *pp.* 1–15

Model Checking via Parametric Bootstraps in Time Series Analysis

By RUEY S. TSAY†

University of Chicago, USA

[Received June 1989. Final revision June 1990]

SUMMARY

This paper uses parametric bootstraps in conjunction with selected functionals such as the spectral density function to derive methods for model checking in time series analysis. The methods proposed emphasize the reproducibilities of the fitted models. They are widely applicable and easy to implement. In particular, they can be used to check special characteristics of the underlying process such as time reversibility and long memory dependence. The paper also addresses the importance of model-building objectives in model checking. Several examples including a wind speed data set for Ireland are used to illustrate the procedures proposed.

Keywords: Diagnostic statistics; Empirical distribution function; Reproducibility; Spectral density function; Stationarity; Time reversibility

1. Introduction

Parametric approaches that attempt to fit parsimonious models to time series data have played an important role in applied time series analysis. The use of such approaches does not necessarily assert the existence of a true model; it merely reflects the belief that a well-chosen model can provide an adequate approximation to the process under study. A critical step in any successful application of parametric approaches is therefore to check the adequacy of a fitted model. For example, diagnostic checking (or model criticism) is regarded as an important step in the iterative modelling procedure of Box and Jenkins (1976). Many methods are available in the literature for checking a fitted time series model. For example, Newbold (1983) provides an excellent survey of many classical goodness-of-fit tests, Chang *et al.* (1988) and Bruce and Martin (1989) address the issues of detecting outliers and interventions, and Tong (1990) discusses many statistics for testing linearity of a time series process.

To assess the validity of a fitted model, we typically report the following:

(a) the mean-squared error of residuals,
(b) the Q-statistics of Ljung and Box (1978),
(c) some residual plots and
(d) some results of outlier detection.

†*Address for correspondence*: Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637, USA.

 0035–9254/92/41001 $2.00

The mean-squared error of residuals is used to measure the fidelity of the model to the data within the data span, the Q-statistic is to check the serial correlations of the residuals, and residual plots and outlier detection are used to spot outliers, non-linearities and possibly variance changes. However, it is well known that the best model with respect to one checking criterion may fare badly with respect to another criterion. For instance, a model with the smallest residual mean-squared error may perform poorly in an out-of-sample forecasting exercise. Consequently, there is a need to specify the objective of data analysis before choosing a checking criterion to assess the adequacy of a fitted model. Without mentioning objectives, reported model checking statistics are meaningless or could be misleading. If the objective of an analysis is to produce reliable forecasts, the out-of-sample prediction may be a criterion to use. In general, if we fear that the fitted model might be vulnerable against a specific alternative, we can use the likelihood ratio or the Lagrange multiplier tests to check the model. For example, Engle (1982) used a Lagrange multiplier test to detect the conditional heteroscedasticity; Keenan (1985), Tsay (1986) and Saikkonen and Luukkonen (1988) used Lagrange multiplier tests or their equivalent to check the non-linearity of a process; Chang *et al.* (1988) and Tsay (1988) used approximate likelihood ratio tests to detect outliers in a time series; Chan and Tong (1986) employed a likelihood ratio test to determine the need for a threshold autoregressive (TAR) model.

The preceding goodness-of-fit tests, however, depend heavily on asymptotic results, and many of them also depend on the normality assumption of the process. The finite sample behaviour of the tests is largely unknown or may be too complicated to be of practical interest. When the distribution of the process (or the associated innovations) is not Gaussian, certain modifications of the existing tests are often needed. For instance, Li and McLeod (1988) considered autoregressive moving average (ARMA) modelling with non-Gaussian innovations and showed that, under certain regularity conditions, the Q-statistic continues to apply provided that the covariance matrix of the residual autocorrelations is properly modified. Such a modification may be cumbersome in some cases, however. Furthermore, there are situations in which the particular feature of interest cannot be characterized by a small number of parameters. Consider, for example, the time reversibility of a univariate time series. It is obvious that the second-order statistics commonly used in data analysis are useless in detecting the existence of such a feature. Any characterization that can capture time irreversibility must involve higher order statistics. A possibility then is to take advantage of higher order spectral density functions. As discussed in Brillinger and Rosenblatt (1967) a strictly stationary time series is time reversible if and only if the imaginary parts of all the higher order spectra are identically zero. However, it is non-trivial to test the zero imaginary parts of higher order spectra in practice. Consequently, further development in time series model checking is needed.

The objective of this paper, therefore, is to provide model checking procedures that are easy to implement, powerful in application and flexible. The flexibility of the procedures proposed stems from the fact that they not only provide an overall evaluation of the fitted model but also can be tailored to meet certain specific needs of the analysis. The ease in implementation is to make use of the available advanced computing facilities, rather than to derive test statistics that are simple in theory. The basic tools of the procedures proposed are parametric bootstraps and some special functionals designed to show the specific features of interest. The basic idea is that a fitted model is adequate if it can successfully reproduce the characteristics of interest. In other words, the paper emphasizes reproducibility of a fitted model in assessing its validity. Suppose, for instance, that we are interested in the generating mechanism of a process. Then, we say that a parametric model is adequate if it can reproduce spectral density functions that are close to the sample spectral density function of the data. Here the spectral density function is the special functional. Since spectral density is a summary of the underlying process, especially when the second-order properties are of major concern, the procedures proposed thus provide an overall evaluation of the fitted model. Also, since the spectral density function allows concentration on a particular frequency, the procedures proposed are therefore flexible and can be tailored to meet the requirement at a given frequency, e.g. checking the fundamental frequency of a periodic time series.

The idea used in this paper is closely related to the Monte Carlo tests of Barnard (1963), Hope (1968) and Besag and Diggle (1977). A difference is that we emphasize model checking with specific objectives instead of significance tests. Some of the graphic displays used are from the concept of 'envelopes' in Ripley (1977).

The paper is organized as follows. Section 2 gives the procedures proposed and discusses some special functionals that are relevant to time series analysis. Section 3 illustrates the procedures by checking several models proposed in the literature for various time series. A remark on computing is given at the end.

2. Procedures

The proposed model checking procedures need

(a) a parametric model and
(b) one or several functionals that can adequately describe the special features of interest.

It is crucial to the success of the procedure proposed that we can use the model to generate sample processes repeatedly. Thus, a parametric model of this paper must consist of

(a) a mathematical form with known parameters and
(b) a known probability distribution for the innovations.

In some cases, starting values of the process may also be important. However, if the process is stationary, the effects of any starting values will be negligible provided that we discard certain data points at the beginning of a generating exercise.

Once the preceding two requirements are satisfied, the parametric model can be linear or non-linear, Gaussian or non-Gaussian and univariate or multivariate. It may also be a regression model with time series errors when the regressors are treated as fixed variables. The basic requirement of the functionals used is that they can adequately describe the special feature of interest. If the special feature of interest can be captured by a parameter, then an obvious functional to use is some sufficient statistic of that parameter. For those situations in which the special feature of interest cannot be easily parameterized, e.g. the time irreversibility mentioned earlier, we may use more than one functional or choose a functional that is simple and easily interpretable. Some useful functionals for time series model checking will be given later. In view of the versatility of possible models and the flexibility in choosing functionals, the checking procedures considered in this paper are widely applicable.

2.1. *The Procedures*

The essence of the proposed model checking procedures is to obtain an empirical distribution of the specified functional via parametric bootstraps. This distribution function then serves as a reference distribution to which the corresponding functional quantity of the data can be compared. More specifically, the fitted model is used repeatedly to generate many samples, each of which has the same number of observations as the original data. The samples generated are then used to construct an empirical distribution of the functional of interest. If the fitted model is adequate in describing the feature of interest, the functional quantity of the original data should be a reasonable point with respect to the empirical distribution. In other words, an adequate model will be able to reproduce the quantity of interest so that that of the data is a regular point of the empirical distribution. In what follows, we give further details of the proposed model checking procedures based on the objectives of a data analysis.

2.2. *Overall Checking Procedure*

The proposed overall assessment of a fitted model for a time series z_t is to use the sample spectral density function of the process (or some filtered process of it) as a functional. For each of the data sets generated, we estimate the spectral density function by using exactly the same method as that used in estimating the sample spectral density function of the observed data, e.g. using the same covariance-lag window with the same truncation point. By so doing, we obtain an empirical distribution of the spectral density function of the fitted model. For each fixed frequency, we can then obtain a $100(1-\alpha)\%$ probability interval of the spectral density function by using the quantiles of the empirical distribution. Suppose that we use the $100(1-\alpha/2)\%$ and $(100\alpha/2)\%$ quantiles at each frequency. Then, an envelope can be obtained for the spectral density function by connecting these quantiles. If the fitted model is adequate at each frequency, the sample spectral density function of the original data should be largely within the envelope. By plotting the envelope and the sample spectral density function of the data jointly, we obtain a graphical display that can be used to assess the overall goodness of fit of the fitted model. This type of envelope was used by Ripley (1977) in conjunction with spatial statistics and, for convenience, they will be referred to as acceptance envelopes. Since (smoothed) sample power spectra at two frequencies are correlated, the probability that the envelope contains the sample spectral density function is usually not $1-\alpha$. Thus the acceptance envelope considered is not a joint $100(1-\alpha)\%$ confidence interval of the sample spectral density function.

To illustrate the idea of an acceptance envelope of a spectral density function, Fig. 1(a) shows the sample spectral density of series A of Box and Jenkins (1976) and an acceptance envelope constructed by using $\alpha=0.10$ and 200 realizations generated from the ARMA(1, 1) model

$$(1-0.9048B)z_t = (1-0.5772B)a_t$$

with a_t independently and identically distributed (IID) as $N(0, 0.0976)$ and B

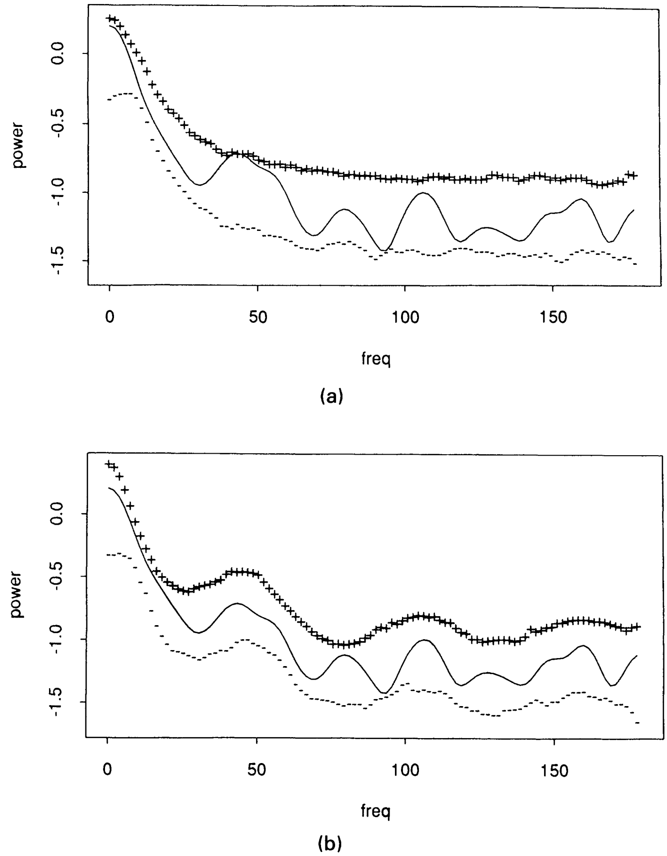


Fig. 1. Acceptance envelopes for series A of Box and Jenkins (1976): (a) ARMA(1, 1) model; (b) AR(7) model

the backshift operator. All the spectral density functions were obtained by using a Parzen window with a truncation point $M=42$. The + line gives the 95-percentile of the power spectrum at each frequency whereas the − line gives the 5-percentile. Fig. 1(b) gives the corresponding results for the AR(7) model

$$(1-0.344B-0.201B^2-0.027B^3-0.016B^4+0.020B^5-0.081B^6-0.191B^7)z_t = 2.73 + a_t$$

where a_t are IID $N(0, 0.0901)$. From the plots, the Gaussian ARMA(1, 1) model appears to be reasonable, except for the frequencies close to 45°. The Gaussian AR(7) model, in contrast, seems to fit the data very well. Note that the constant term is dropped in the simulation, because it has no effects on the spectral density estimation. Also, for each realization of n observations we generated $n+400$ data points with zero starting values, but only the last n data points were used in spectral density estimation.

2.3. *Local Assessment*

In some situations, special emphasis is placed on a particular frequency of the process. For instance, in analysing a periodic time series, we may emphasize the fundamental frequencies rather than all the frequencies. The comparison between sample spectral density functions of the model and the data appears to be well suited for such analyses. Suppose that we are interested in the periodicity of a given time series that corresponds to the frequency ω. In this case, by concentrating on the power at ω, the bootstrapping procedure of the preceding paragraph gives an empirical distribution of the power spectrum at ω. The local assessment proposed is then to compute the p-value of the power spectrum of the data at frequency ω with respect to the empirical distribution.

2.4. *Time Reversibility*

A characterization of stationary linear Gaussian time series is that the process is reversible in time. Thus it is of common interest to check the time reversibility of a time series. As mentioned before, it is impossible to characterize time reversibility by a single parameter. Indeed, many statistics have been considered in the literature to capture time reversibility. Weiss (1975) proposed to take advantage of the symmetry of the distribution of a time reversible process z_t. Another possibility to check time reversibility is to use the statistics

$$R_k = E(z_t^2 z_{t-k} - z_t z_{t-k}^2) = E(z_t z_{t+k}^2 - z_t z_{t-k}^2), \qquad k > 0,$$

where the second equality holds only under stationarity of z_t, but it emphasizes that R_k depends on the difference in covariances expressed forwards in time and backwards in time. Under the hypothesis that z_t is reversible in time, the expectation of R_k is zero.

In this paper, we employ another statistic for describing time reversibility which is particularly useful for periodic time series. Define a dichotomous process y_t by

$$y_t = \begin{cases} 1 & \text{if } z_t - z_{t-1} \geqslant 0, \\ -1 & \text{if } z_t - z_{t-1} < 0. \end{cases} \tag{1}$$

Consider the pair (z_{t-1}, z_t) of the original process and the pair (z_t, z_{t-1}) of the reverse process. If the process is reversible in time, the probability distributions of these two pairs are identical. In other words, z_t and z_{t-1} are exchangeable. Consequently, the following property holds.

Property 1. Suppose that z_t is a strictly stationary and time reversible process. Define y_t by process (1). Then $P(y_t = 1) = P(y_t = -1) = \frac{1}{2}$ for all t.

Let P_n and N_n be respectively the number of 1s and -1s in $y_2, \ldots, y_n$, and define $RA_n = P_n/N_n$. By property 1, the law of large numbers and Slutsky's theorem (Bickel and Doksum, 1977) if the process z_t is ergodic and time reversible, then RA_n converges in probability to unity as $n \to \infty$. For fixed and sufficiently large n, RA_n is a random variable with mean around unity. Thus any significant departure of RA_n from unity indicates time irreversibility. Consequently, RA_n can serve as a functional for checking time reversibility of z_t. For convenience, we refer to RA_n as the positive/negative (PN) ratio of z_t.

2.5. *Long Memory Dependence*

In recent years, there has been a growing interest in studying the long memory dependence of a time series and fractional difference is regarded as an efficient way to model such a process. The simplest fractionally differenced model is

$$(1-B)^d z_t = a_t \tag{2}$$

where $0 \leqslant d \leqslant 0.5$ is the fractional difference. See Granger and Joyeux (1980), Hosking (1981) and Geweke and Porter-Hudak (1983). Although the usual model checking procedures such as the Q-statistic of Ljung and Box (1978) apply in diagnosing a fitted fractionally differenced model, it seems that a more appropriate checking procedure is to evaluate the ability of a fitted model in reproducing the long memory dependence. Therefore, we use as a functional the statistic $\hat{d}$ defined by

$$\hat{d} = -\sum_{j=1}^{m} (x_j-\bar{x})(y_j-\bar{y}) \Bigg/ \sum_{j=1}^{m} (x_j-\bar{x})^2 \tag{3}$$

where m is a prespecified positive integer, $\bar{y}$ and $\bar{x}$ are the averages of y_j and x_j respectively, $y_j = \ln I_n(\omega_j)$, $x_j = \ln\{|1-\exp(-\mathrm{i}\omega_j)|\}$, where $I_n(\omega_j)$ is the sample periodogram of z_t at the frequency $\omega_j = 2\pi j/n$, and n is the sample size of the original data. Geweke and Porter-Hudak (1983) showed that under some regularity conditions $\hat{d}$ is asymptotically normal with mean d and variance $\pi^2/6\Sigma_{j=1}^{m}(x_j-\bar{x})^2$.

2.6. *Stability*

A characteristic that is difficult to quantify in time series analysis is the degree of non-stationarity (or stability) of a process. This is related to the problem of determining the order of differencing for which many tests have been proposed with their critical values tabulated, e.g. Fuller (1976). However, two random walk processes may behave very differently, and there are situations in which the distinction between stationarity and non-stationarity of a process is at best an analyst's subjective opinion. Thus there is a need to quantify the degree of stability of a time series. In this paper, we use the unit root case and ARMA models to discuss a measure that can quantify the degree of non-stationarity of a time series. Extensions to other non-stationary characteristic roots on the unit circle are trivial.

Suppose that we are interested in measuring the degree of non-stationarity of a time series z_t with respect to a unit root factor $(1-B)^h$, where h is a predetermined positive integer. Suppose also that $y_t = (1-B)^h z_t$ satisfies the model

$$\phi(B)y_t = c + \theta(B)a_t \tag{4}$$

where c is a constant, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ have no common factors and have all their zeros on or outside the unit circle. Define $S(h)$ as the degree of non-stationarity of z_t with respect to $(1-B)^h$ by

$$S(h) = \begin{cases} |\theta(1)/\phi(1)|\,\sigma_a & \text{if } \phi(1) \neq 0, \\ \infty & \text{otherwise,} \end{cases} \tag{5}$$

where σ_a is the standard deviation of the innovational process a_t. Such a definition has several interpretations. First, if $y_t = (1-B)^h z_t$ is stationary, then it is easy to see that

$$S^2(h) = 2\pi\, f(0)$$

with

$$f(\omega) = \frac{\sigma_a^2}{2\pi} \sum_{\nu=-\infty}^{\infty} \rho_\nu \exp(-\mathrm{i}\nu\omega) \tag{6}$$

being the spectral density function of y_t, where ρ_ν is the lag ν autocorrelation of y_t. Second, it can be shown that $S(h)$ is related directly to the standard deviations of the multistep-ahead forecasts of z_t if $(1-B)^h z_t$ is stationary and $h > 0$. More specifically, let $\hat{e}(l) = z_t - \hat{z}_t(l)$, where $\hat{z}_t(l)$ is the l-step ahead forecast of z_t at time index t. Then, we have

$$\lim_{l\to\infty} [\mathrm{var}\{\hat{e}(l)\}/l^{2h-1}] = c(h, \phi, \theta)\, S^2(h),$$

where $c(h, \phi, \theta)$ is a constant depending on h, ϕ and θ of equation (4). In general, a small $S(h)$ means that z_t is stable with respect to the non-stationary factor $(1-B)^h$. The following properties of $S(h)$ are easy to prove and insightful in understanding the measure.

Property 2. Suppose that z_t is a time series process such that $(1-B)^d z_t$ is stationary. Also, assume that the non-stationary factor of interest is $(1-B)^h$. Then

(a) $S(h) = 0$ if $d < h$,
(b) $S(h) = 0$ if z_t is deterministic, i.e. $\sigma_\mathrm{a}^2 = 0$,
(c) $S(h) = \infty$ if $d > h$ and
(d) $S(h) < \infty$ if $d = h$.

Obviously, we can use the stability measure $S(h)$ of definition (5) as a functional when we are interested in determining the order of difference of a time series.

3. Applications

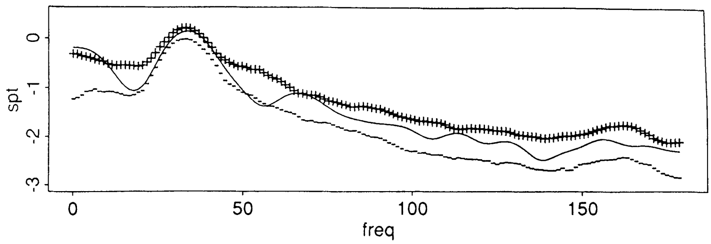
In this section, we illustrate the proposed model checking procedures of Section 2 by analysing several models for three data sets including the wind speed data of Haslett and Raftery (1989) for Ireland.

3.1. *Example 1*

Consider the much investigated annual Wolf sunspot numbers from 1700 to 1979 and suppose that the objective of modelling is to describe the underlying mechanism of the data. If Gaussian linear processes were entertained, an AR(9) model appears to be the choice of most model selection criteria available in the literature. The estimated model is

$$\begin{aligned} z_t = 6.96 &+ 1.21z_{t-1} - 0.45z_{t-2} - 0.17z_{t-3} + 0.20z_{t-4} - 0.13z_{t-5} \\ &+ 0.03z_{t-6} + 0.01z_{t-7} - 0.03z_{t-8} + 0.21z_{t-9} + a_t \end{aligned} \tag{7}$$

with $\sigma_a^2 = 221.24$. However, if we used TAR models and modified model (9.2) of Tong and Lim (1980) to account for the full data span, we obtained the model



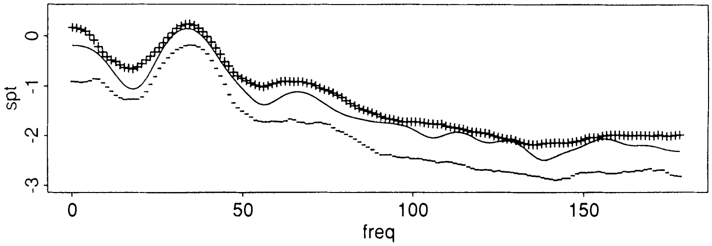


Fig. 2. Acceptance envelopes for annual sunspot numbers: (a) TAR model; (b) AR(9) model

$$z_t = \begin{cases} 10.88 + 1.869z_{t-1} - 1.556z_{t-2} + 0.086z_{t-3} + 0.326z_{t-4} + a_{1t}, & \text{if } z_{t-3} \leqslant 32.3, \\ 8.726 + 0.679z_{t-1} + 0.064z_{t-2} - 0.217z_{t-3} + 0.044z_{t-4} & \\ - 0.118z_{t-5} - 0.005z_{t-6} + 0.192z_{t-7} - 0.285z_{t-8} + 0.242z_{t-9} & \\ - 0.123z_{t-10} + 0.246z_{t-11} + 2a_{2t}, & \text{otherwise}, \end{cases} \quad (8)$$

where a_{1t} and a_{2t} are independent Gaussian random variates with mean zero and variances 275.7 and 82.9 respectively.

To evaluate models (7) and (8), we consider two functionals. The first functional is the sample spectral density function and the second is the PN ratio. The second functional is used because the sunspot data clearly exhibit an asymmetric pattern in ascending and descending times in each cyle. Figs 2(a) and 2(b) show the acceptance envelopes of models (8) and (7) respectively along with the sample spectral density function of the data. All spectral densities are in the logarithmic scale. The envelopes were obtained by using 300 realizations each with 280 observations. A truncation point $M = 44$ and $\alpha = 0.02$ were used in all spectral density estimations. From the plots, it is seen that in terms of the second-order properties the two models are reasonable. The TAR model seems to have a problem at the lower frequency, but it describes the spectrum better at the higher frequency.

Fig. 3 shows the empirical distribution functions of the PN ratio of models (8) and (7) with the shaded function corresponding to the linear AR(9) model. The difference between the two models is clearly seen. The PN ratio of the sunspot data is 0.743 from which we can compute the p-values, i.e. $\Pr(RA_n \leqslant 0.743)$. For the linear AR(9) model, the p-value is less than 0.003 whereas that of the TAR model is 0.20. Thus the non-linear model can reproduce the asymmetric pattern of the data, but the linear AR(9) model cannot.

Li and McLeod (1988) used the data span from 1770 to 1955 and log-normal innovations to obtain the model

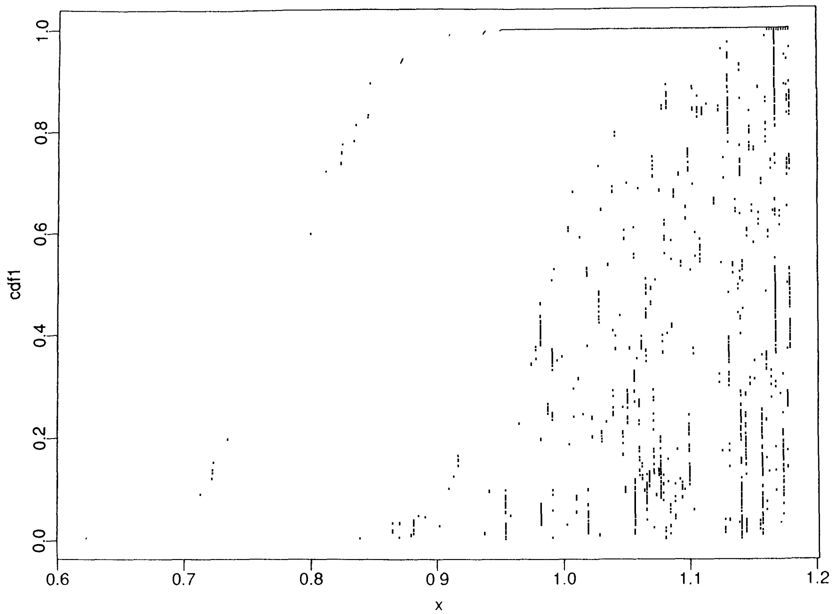


Fig. 3. Empirical distribution functions of the PN ratio for annual sunspot numbers: TAR and AR(9) models

$$z_t = 1.6759z_{t-1} - 0.7840z_{t-2} + b_t \tag{9}$$

where b_t is log-normally distributed with estimated mean 13.88 and variance 153.39. They used the residual mean-squared error to suggest that the above log-normal AR(2) model is comparable with other models available in the literature for the sunspot data. Fig. 4 shows the acceptance envelope of the spectral density function of model (9) with $\alpha = 0.02$ and $M = 44$. Again, 300 realizations were used and each realization contains 280 observations. From the plot, the adequacy of the fitted log-normal AR(2) model is questionable. In particular, the model does not have a spectral peak at a frequency around 11 years!

3.2. *Example 2*

Haslett and Raftery (1989) used fractional difference to model the long memory dependence of mean daily wind speeds at 12 synoptic meteorological stations in Ireland. The data were from 1961 to 1978 giving 6574 observations for each series. The basic model of Haslett and Raftery for daily wind velocities, wind speeds after removing seasonal components, is

$$(1 - 0.010B + 0.063B^2)(1 - B)^{0.328}z_t = a_t \tag{10}$$

with a_t IID $N(0, 0.246)$. Since fractional difference is the key feature emphasized in the analysis, we used the statistic $\hat{d}$ of equation (3) as a functional to check the adequacy of model (10).

Owing to the long memory dependence of the fitted model (10), each of the parametric bootstrap samples was generated by the following steps.

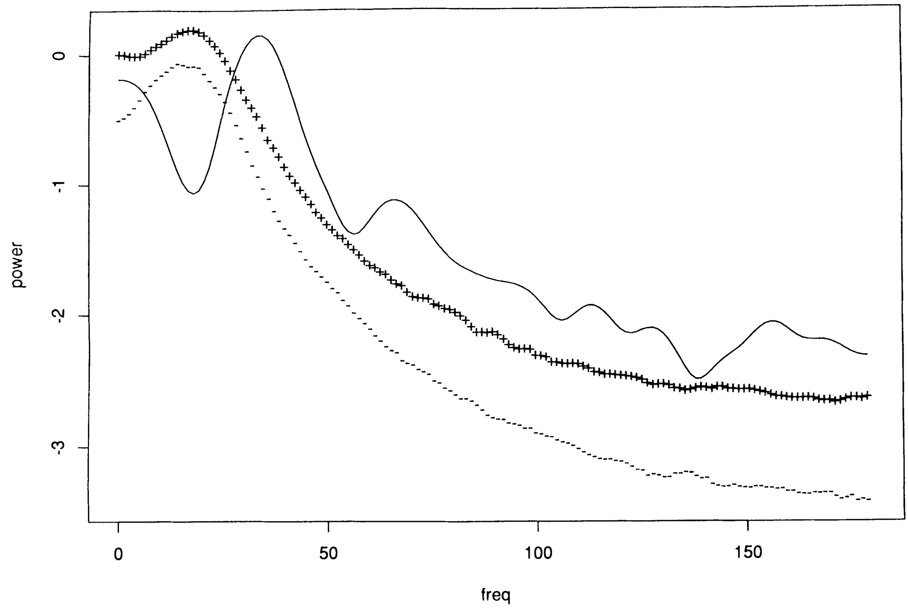


Fig. 4. Acceptance envelope for annual sunspot numbers based on a log-normal AR(2) model

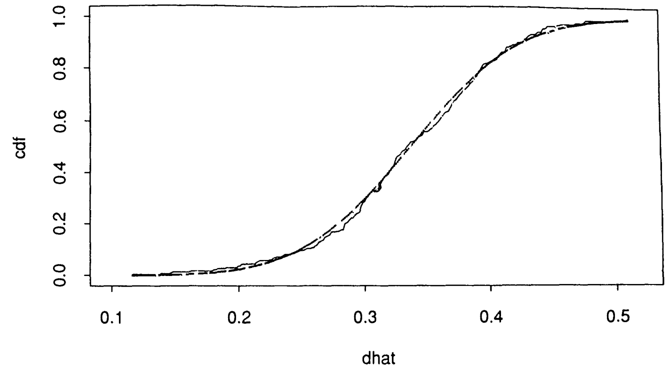
(a) Generate 9874 observations from the AR(2) process $(1-0.010B+0.063B^2)y_t = a_t$ with a_t being IID $N(0, 0.246)$ and $y_0 = y_{-1} = 0$.
(b) For $t=3301$ to $t=9874$, use the first 3000 terms of the binomial expansion of $(1-B)^{-d}$ with $d=0.328$ to generate z_t from y_t, i.e.

$$z_t = \sum_{k=0}^{3000} \psi_k y_{t-k}$$

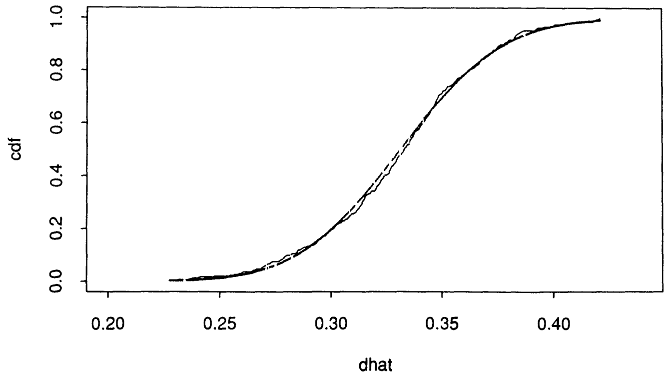
with $\psi_k = (k+d-1)!/k!(d-1)!$.

See equation (3.1) of Hosking (1981). Here 3000 terms were used to obtain an accurate approximation of the fractional difference $(1-B)^d$, and the first 300 data points of y_t were discarded to reduce the effect of using $y_0 = y_{-1} = 0$.

Once a sample of 6754 data points had been generated, the estimate of the fractional difference d was done by using equation (3) with $m=100$ and $m=300$. This exercise was carried out in 300 replications. Fig. 5 gives the empirical cumulative density functions (CDFs) of the estimates of fractional difference (the full lines) as well as the corresponding CDFs of Gaussian distributions (the broken lines). The Gaussian CDFs were obtained by using the sample mean and standard deviation of the estimates. On the basis of the plots, the estimates follow closely the theory given in Geweke and Porter-Hudak (1983). The asymptotic mean and standard deviation of $\hat{d}$ for $m=100$ are 0.328 and 0.0695 respectively. The results of the parametric bootstraps are 0.337 and 0.0689. For $m=300$, the mean and standard deviation of the asymptotic theory are 0.328 and 0.0384 whereas those of the empirical results are 0.331 and 0.0368. Thus, Fig. 5 shows that the empirical distributions of $\hat{d}$ are reasonable.



(a)



(b)

Fig. 5. Empirical distribution functions of fractional difference for Ireland's wind speed data: (a) $m = 100$; (b) $m = 300$

Table 1 gives the estimates $\hat{d}$ and the corresponding p-values for the wind velocities by station. The p-values are the probabilities $\Pr(d \leqslant \hat{d})$ based on the empirical distributions of Fig. 5. The results in the table give little support for long memory dependence with $d = 0.328$ for the wind velocities. In particular, the $\hat{d}$ for all 12 stations are less than the minimum of the estimates of bootstrap samples when $m = 300$.

3.3. *Example 3*

Consider the mid-year resident US population from 1930 to 1982. This series has been analysed by Dickey *et al.* (1986) and Cressie (1988). To achieve stationarity, the need for the first difference $1 - B$ is obvious. However, the need for a second difference, i.e. taking $(1 - B)^2$, is not clear. On the basis respectively of unit root tests and a graphical procedure, the preceding two papers support the second difference. Furthermore, Dickey *et al.* (1986) concluded that for this data set the second difference $(1 - B)^2 z_t$ is white noise, i.e. $(1 - B)^2 z_t = a_t$. In this paper, since we focus on checking the second difference and the white noise model for the second-differenced data, we shall use $y_t = (1 - B) z_t$ as the observed data. In other words, we employ the

TABLE 1
Estimates of fractional differences for Ireland's daily wind velocities by station†

Station	$m=100$ $\hat{d}$	$Pr(d \leqslant \hat{d})$	$m=300$ $\hat{d}$	$Pr(d \leqslant \hat{d})$
Roche's Point	0.163	0.016	0.058	<0.003
Valentia	0.142	0.007	0.114	<0.003
Rosslare	0.060	<0.003	0.076	<0.003
Kilkenny	0.203	0.043	0.084	<0.003
Shannon	0.219	0.057	0.104	<0.003
Birr	0.309	0.330	0.160	<0.003
Dublin	0.198	0.030	0.155	<0.003
Claremorris	0.379	0.710	0.202	<0.003
Mullingar	0.169	0.016	0.097	<0.003
Clones	0.293	0.233	0.149	<0.003
Belmullet	0.164	0.016	0.127	<0.003
Malin Head	0.191	0.030	0.096	<0.003

†<0.003 denotes that the estimate is less than the minimum of the parametric bootstrap estimates.

random walk model

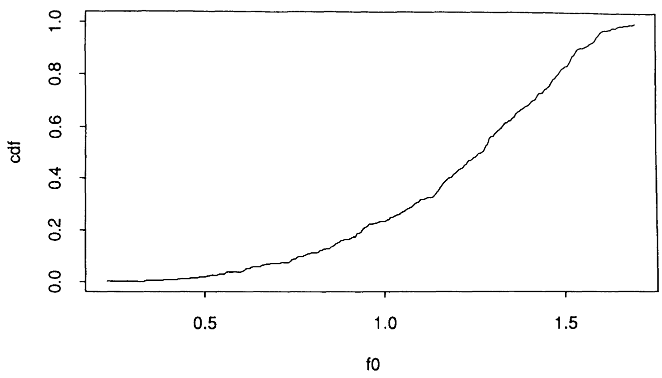
$$(1-B)y_t = a_t \tag{11}$$

with $\sigma_a^2 = 75005$, for y_t to check the need for a second difference of z_t and to assess the adequacy of the model $(1-B)^2 z_t = a_t$. The sample spectral density function and the degree of stability $S(1)$ in definition (5) of y_t were used as functionals.

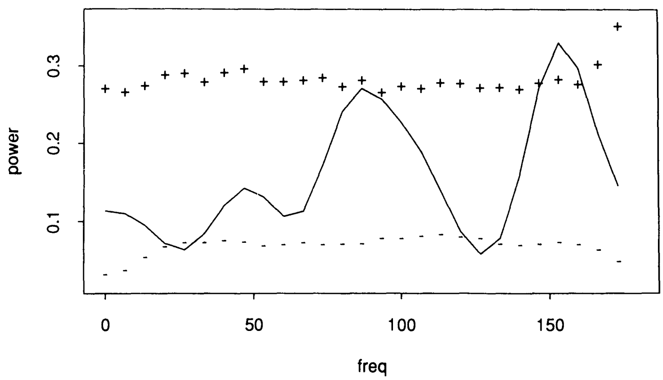
Fig. 6(a) gives the empirical CDF of the power spectrum of y_t at zero frequency and Fig. 6(b) shows the acceptance envelope of the spectral density function of $(1-B)y_t = a_t$ with $\alpha=0.05$ and $M=20$ and the sample spectrum of the twice-differenced original data, $(1-B)^2 z_t$. From the acceptance envelope, the suggested model of Dickey *et al.* (1986) is questionable, for several frequencies having power outside the envelope. However, the sample power spectrum of $(1-B)z_t$ at zero frequency is 1.5467 which is a reasonable point with respect to the empirical CDF of Fig. 6(a). Our model checking procedure thus supports the need for a second difference. Since the process is highly non-stationary, for each bootstrap sample 600 data points were discarded at the beginning of the generating exercise.

4. Concluding Remarks

In this paper we have considered model checking in time series analysis by using parametric bootstraps and some pivotal functionals. Real examples were used to illustrate the flexibility and power of the proposed model checking procedures. Since computing will be cheaper and easier as time passes, we believe that the procedures proposed will be widely available in the future. For the examples used in the paper, the central processor unit (CPU) times for examples 1 and 3 were trivial. In contrast example 2 required approximately 16 h of CPU time on a Microvax Workstation 3200. This is due to the large sample size $n=6574$ and the long binomial expansion used to generate the parametric bootstrap samples.



(a)



(b)

Fig. 6. Model checking for US mid-year population: (a) empirical CDF of the power spectrum at zero frequency; (b) acceptance envelope for the sample spectral density function

Acknowledgements

The author wishes to thank Adrian Raftery for providing the wind speed data for Ireland and the referees and Rong Chen for useful comments. This research is supported in part by the National Science Foundation and by the William S. Fishman Research Fund, Graduate School of Business, the University of Chicago.

References

Barnard, G. A. (1963) Discussion on The spectral analysis of point processes (by M. S. Bartlett). *J. R. Statist. Soc.* B, **25**, 294.

Besag, J. and Diggle, P. J. (1977) Simple Monte Carlo tests for spatial pattern. *Appl. Statist.*, **26**, 327–333.

Bickel, P. J. and Doksum, K. A. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*, p. 461. San Francisco: Holden-Day.

Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis: Forecasting and Control*, 2nd edn. San Francisco: Holden-Day.

Brillinger, D. R. and Rosenblatt, M. (1967) Computation and interpretation of *k*th order spectra. In *Spectral Analysis of Time Series* (ed. B. Harris), pp. 153–188. New York: Wiley.

Bruce, A. G. and Martin, R. D. (1989) Leave-*k*-out diagnostics for time series (with discussion). *J. R. Statist. Soc.* B, **51**, 363–424.

Chan, W. S. and Tong, H. (1986) On tests for non-linearity in time series analysis. *J. Forecast.*, **5**, 217–228.

Chang, I., Tiao, G. C. and Chen, C. (1988) Estimation of time series parameters in the presence of outliers. *Technometrics*, **30**, 193–204.

Cressie, N. (1988) A graphical procedure for determining nonstationarity in time series. *J. Am. Statist. Ass.*, **83**, 1108–1116.

Dickey, D. A., Bell, W. R. and Miller, R. B. (1986) Unit roots in time series models: tests and implications. *Am. Statistn*, **40**, 12–26.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, **50**, 987–1008.

Fuller, W. A. (1976) *Introduction to Statistical Time Series*. New York: Wiley.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.*, **4**, 221–238.

Granger, C. W. J. and Joyeux, R. (1980) An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.*, **1**, 15–29.

Haslett, J. and Raftery, A. E. (1989) Space–time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Appl. Statist.*, **38**, 1–50.

Hope, A. C. A. (1968) A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc.* B, **30**, 582–598.

Hosking, J. R. M. (1981) Fractional differencing. *Biometrika*, **68**, 165–176.

Keenan, D. M. (1985) A Tukey non-additivity-type test for time series non-linearity. *Biometrika*, **72**, 39–44.

Li, W. K. and McLeod, A. I. (1988) ARMA modelling with non-Gaussian innovations. *J. Time Ser. Anal.*, **9**, 155–168.

Ljung, G. M. and Box, G. E. P. (1978) On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

Newbold, P. (1983) Model checking in time series analysis. In *Applied Time Series Analysis of Economic Data* (ed. A. Zellner), pp. 138–143. Washington DC: US Bureau of the Census.

Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B, **39**, 172–212.

Saikkonen, P. and Luukkonen, R. (1988) Lagrange multiplier tests for testing non-linearities in time series models. *Scand. J. Statist.*, **15**, 55–68.

Tong, H. (1990) *Non-linear Time Series Analysis: a Dynamical System Approach*. Oxford: Oxford University Press.

Tong, H. and Lim, K. S. (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *J. R. Statist. Soc.* B, **42**, 245–292.

Tsay, R. S. (1986) Non-linearity tests for time series. *Biometrika*, **73**, 461–466.

——(1988) Outliers, level-shifts and variance changes in time series. *J. Forecast.*, **7**, 1–20.

Weiss, G. (1975) Time-reversibility of linear stochastic processes. *J. Appl. Probab.*, **12**, 831–836.