Physica A 389 (2010) 1601–1606

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A note on the sum of the sample autocorrelation function

Hossein Hassani *

Statistics Group, Cardiff School of Mathematics, Cardiff University, CF24 4AG, UK
2-Statistical Research and Training Center, Tehran, 1413717911, Iran

ARTICLE INFO

Article history:
Received 23 October 2009
Received in revised form 16 December 2009
Available online 4 January 2010

Keywords:
Sample autocorrelation function
Stationary process
Diagnostic check
Approximate test for autocorrelation

ABSTRACT

It is shown that the sum of the sample autocorrelation function at lag $h \geq 1$ is always $-\frac{1}{2}$ for any stationary time series with arbitrary length $T \geq 2$ (Hassani, 2009 [1]). In this paper, the distribution of a set of the sample autocorrelation function using the properties of this quantity is considered. It is found that the distribution of a set of the sample autocorrelation estimates is not independent and identically distributed. This finding implies that the result of diagnostic check and model building using the traditional assumption of *iid* can be quite misleading.

1. Introduction

In time series analysis, the autocorrelation function and partial autocorrelation function play important roles in data analysis and lag identification in an autoregressive model. The distribution of a set of the sample autocorrelation estimates, scaled by the square root of the sample size, found in many standard time series textbooks and software, is assumed to be independent and identically distributed. Therefore, the result of diagnostic check and model building is misleading if this assumption is not met. Hassani [1] considered the sum of the sample autocorrelation function, S_{acf}, of any linear stationary process. Using the properties of S_{acf}, he showed that the sample spectral density of a stationary process fluctuates radically about the theoretical spectral density. In the present study, I use the properties of S_{acf} in examining the distribution of the sample autocorrelation function, and approximate test for autocorrelation.

The autocovariance function of a wide sense stationary process $\{Y_t, t \in \mathbb{N}\}$ at lag h is:

$$R(h) = E[(Y_{t+h} - \mu_Y)(Y_t - \mu_Y)], \quad h \in \mathbb{Z} \tag{1}$$

where E is the expected value operator, μ_Y is the expected value of the variable Y. In practical problems we only have a set of data $Y_T = (y_1, \ldots, y_T)$; the following estimator can be considered as an estimate of $R(h)$:

$$\tilde{R}(h) = \begin{cases} \dfrac{\sum_{t=1}^{T-|h|} (y_{t+|h|} - \bar{y})(y_t - \bar{y})}{T - |h|} & h = 0, \pm 1, \ldots, \pm(T-1) \\ 0 & |h| \geq T, \end{cases} \tag{2}$$

where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ is the sample mean which is an unbiased estimator of μ. If we ignore the effect of estimating μ by $\bar{y}$ (that is if we replace $\bar{y}$ by μ in (2)) then it is easy to show that $\tilde{R}(h)$ is an unbiased estimator of $R(h)$. In the general case it can be shown that $\tilde{R}(h)$ is an asymptotically unbiased estimator [2].

* Corresponding author. Tel.: +44 029 2087 4811; fax: +44 029 2087 4199.
E-mail address: hassanih@cf.ac.uk.

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.12.050

There is an alternative estimate of $R(h)$ which suggested by some authors (for example Refs. [3,4]):

$$\hat{\gamma}(h) = \begin{cases} \dfrac{\sum_{t=1}^{T-|h|} (y_{t+|h|} - \bar{y})(y_t - \bar{y})}{T} & h = 0, \pm 1, \dots, \pm(T-1) \\ 0 & |h| \geq T, \end{cases} \tag{3}$$

$\hat{\gamma}(h)$ is biased on the use of the divisor T rather than $T - |h|$ and also has larger bias than $\tilde{R}(h)$. Note that the following relationship holds between $\tilde{R}(h)$ and $\hat{\gamma}(h)$:

$$\frac{\tilde{R}(h)}{\hat{\gamma}(h)} = \frac{T}{T - |h|} \quad h = 0, \pm 1, \dots, \pm(T-1). \tag{4}$$

It has been asserted that, in general, $\hat{\gamma}(h)$ has smaller mean squared error than $\tilde{R}(h)$. For example Ref. [3] has shown that for a particular AR(1) process, $\hat{\gamma}(h)$ has smaller mean squared error for all $|h| > 0$, the difference increasing significantly as h increases. It should be noted that $\hat{\gamma}(h)$ is more popular than $\tilde{R}(h)$ among time series analysts and they prefer to use $\hat{\gamma}(h)$ (see, for example, Refs. [5,2,6,7] among others). Furthermore, most standard statistical packages e.g. Minitab, SPSS, Eviews, Stata, Splus, R and SAS use $\hat{\gamma}(h)$ rather than $\tilde{R}(h)$.

Priestley [2] discussed in detail the advantages of using $\hat{\gamma}(h)$ rather than $\tilde{R}(h)$ from several points of view. The main point why $\hat{\gamma}(h)$ is preferred to $\tilde{R}(h)$ is that $\hat{\gamma}(h)$ has positive semi-definite property while $\tilde{R}(h)$ does not, this is an important issue in the spectral domain estimation. Therefore, we consider $\hat{\gamma}(h)$ as an estimator of $R(h)$.

The autocorrelation function, ACF, is given by

$$\rho(h) = \frac{R(h)}{R(0)} \quad h \in \mathbb{Z}, \tag{5}$$

and an estimate of $\rho(h)$ is:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad h = 0, \pm 1, \dots, \pm(T-1). \tag{6}$$

The sequence $\hat{\rho}(h)$ is positive semi-definite (this follows from the property of $\hat{\gamma}(h)$). But this does not necessarily hold for sequence $\tilde{\rho}(h) = \frac{\tilde{R}(h)}{\tilde{R}(0)}$. Moreover, it is easy to show that $|\hat{\rho}(h)| \leqslant 1$ for all h and, again, this property does not necessarily hold for $\tilde{\rho}(h)$ [2, p. 331]. In the rest of the paper, we only consider ACF at lag $h > 0$ since $\rho(h) = \rho(-h)$ and also use $\hat{\rho}(h)$ as its estimator. Note also that $\rho(0) = 1$. Let us now consider the sum of sample autocorrelation function.

Theorem 1. *Sum of the sample autocorrelation function, S_{acf}, at lag $h \geq 1$ is always $-\frac{1}{2}$ for any stationary time series with arbitrary length $T \geq 2$; that is $S_{acf} = \sum_{h=1}^{T-1} \hat{\rho}(h) = -\frac{1}{2}$.*

Proof (*See Ref. [1]*). The S_{acf} has the following properties:

1. It does not depend on the time series length T; $S_{acf} = -\frac{1}{2}$ for $T \geq 2$.
2. The value of S_{acf} is equal to $-\frac{1}{2}$ for any stationary time series. Thus, for example, S_{acf} for ARMA (p, q) of any order (p, q) is equal to a Gaussian white noise process and both are equal to $-\frac{1}{2}$.
3. The values of $\hat{\rho}(h)$ are linearly dependent:

$$\hat{\rho}(i) = \frac{-1}{2} - \sum_{j \neq i = 1}^{T-1} \hat{\rho}(j) \quad i = 1, \dots, T-1. \tag{7}$$

4. There is at least one negative $\hat{\rho}(h)$ for any stationary time series even for AR(p) with positive ACF. □

2. Approximate test for autocorrelation

Let $\{Y_t, t \in \mathbb{N}\}$ be a real valued wide sense stationary stochastic process with autocorrelation function $\rho(.)$ and zero expectation. In the following we consider asymptotic distribution of $\hat{\rho}$. In doing so, we mainly follow Ref. [2]. Let us consider the following theorem [8].

Theorem 2. *Let Y_t be a general linear process of the form*

$$Y_t = \mu + \sum_{u=-\infty}^{\infty} g_u \varepsilon_{t-u}, \tag{8}$$