

ON THE AUTOCORRELATION FUNCTION OF A TRENDED SERIES

C. MAR MOLINERO
UNIV. OF SOUTHAMPTON

Equations are derived for the autocorrelation function of a trended series. The special case of a linear trend is analysed in detail. It is shown that the zero of the autocorrelation function of a trended series is, in general, only dependent on the length of the series. This result is valid for stochastic and deterministic trends.

Keywords: TIME SERIES, ARIMA MODELS, NON-STATIONARY SERIES, AUTOCORRELATION FUNCTIONS.

1. INTRODUCTION.

When analysing a time series it is common practice to use a computer package to calculate the sample autocorrelation function (acf). First the acf of the original series is computed, later the series is transformed, usually by differencing, and the acf of the transformed series is computed. Sample acf are then used to identify a particular stochastic process that will adequately describe the properties of the series.

Series that exhibit a well defined trend show a well defined pattern in the acf. Autocorrelations at low lags are very high, and decline very slowly as the lag increases; they become zero eventually and continue decreasing as the lag goes on increasing as in Figure 1.

This pattern can also be observed in the sample autocorrelation function of some stationary series, thus the model that correctly describes the behaviour of a time series may be difficult to identify. Wichern /5/ claimed that on the basis of the sample correlogram it might be impossible to discriminate between an ARIMA (0,1,1) process and an ARIMA (1,0,1) one. Anderson /1/ suggested that a test based on the point at which the sample autocorrelations of a non-stationary process change from being positive to being negative

might be of use in this situation. He later produced simulation results to explore the power of such a test (Anderson and Gooijer /2/). Anderson's results are, however, limited to the case in which the autoregressive part of the stochastic process contains unit roots. Tests of non-stationarity that relax the unit root assumption are available, see Fuller /4/, but do not seem to be generally used. This paper extends Anderson's results to the general case of a non-stationary series.

2. DERIVATION OF RESULTS.

Assume that the true process is

$$X_t = T_t + \epsilon_t \quad (1)$$

where T_t is a deterministic time trend and ϵ_t is a stationary process. It is further assumed that

$$E[\epsilon_t] = 0 \quad E[T_t \epsilon_{t+k}] = 0 \quad \text{for all } k \quad (2)$$

Sample autocorrelations are computed using

$$r_k = \frac{C_k}{C_0} \quad (3)$$

where C_k is the autocovariance at lag k

- Cecilio Mar Molinero - The University of Southampton - Department of Accounting and Management Science. Southampton, SO9 5NH.

- Article rebut el juny de 1985.

that is computed using

$$C_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (4)$$

and

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \quad (5)$$

Equation (4) divides by n-k instead of the usual formulation that uses only n. The use of n-k considerably simplifies the calculations. A correction factor can be applied later if the more standard formulation is preferred.

Manipulation of equation (4) produces

$$(n-k) C_k = \sum_{t=1}^{n-k} X_t X_{t+k} - (n+k) \bar{X}^2 + \bar{X}(S_I + S_F) \quad (6)$$

where

$$S_I = \sum_{t=1}^k X_t \quad (7)$$

$$S_F = \sum_{t=n-k+1}^n X_t \quad (8)$$

Using equation (1) to substitute for x_t in equation (6) the following expression is derived

$$(n-k)C_k = \sum_{t=1}^{n-k} T_t T_{t+k} + \sum_{t=1}^{n-k} \epsilon_t \epsilon_{t+k} - (n+k)\bar{T}^2 + \bar{T}(T_I + T_F) + A \quad (9)$$

where

$$A = \sum_{t=1}^{n-k} (\epsilon_t T_{t+k} + \epsilon_{t+k} T_t) - (k+n)\bar{\epsilon}(\bar{\epsilon} + 2\bar{T}) + \bar{\epsilon}(\bar{T}_I + \bar{T}_F) + \bar{T}(\sum_{t=1}^k \epsilon_t + \sum_{t=n-k+1}^n \epsilon_t) + \bar{\epsilon}(\sum_{t=1}^k \epsilon_t + \sum_{t=n-k+1}^n \epsilon_t) \quad (10)$$

$$\bar{T} = \frac{1}{n} \sum_{t=1}^n T_t \quad (11)$$

$$\bar{\epsilon} = \frac{1}{n} \sum_{t=1}^n \epsilon_t \quad (12)$$

$$T_I = \sum_{t=1}^k T_t \quad (13)$$

$$T_F = \sum_{t=n-k+1}^n T_t \quad (14)$$

Using

$$\text{cov}(T_t T_{t+k}) = \frac{1}{n-k} \sum_{t=1}^{n-k} T_t T_{t+k} - \bar{T}_t \bar{T}_{t+k} \quad (15)$$

where

$$\bar{T}_t = \frac{1}{n-k} \sum_{t=1}^{n-k} T_t \quad (16)$$

$$\bar{T}_{t+k} = \frac{1}{n-k} \sum_{t=k+1}^n T_t \quad (17)$$

and the relationships

$$T_F = n\bar{T} - (n-k) \bar{T}_t \quad (18)$$

$$T_I = n\bar{T} - (n-k) \bar{T}_{t+k} \quad (19)$$

Equation (9) can be modified to read

$$C_k = \frac{1}{n-k} \sum_{t=1}^{n-k} \epsilon_t \epsilon_{t+k} + \text{cov}(T_t T_{t+k}) + (\bar{T} - \bar{T}_t)(\bar{T} - \bar{T}_{t+k}) + \frac{1}{n-k} A \quad (20)$$

The first term in expression (20) is nothing else than the sample autocovariance of the series ϵ_t at lag k, if we call it $c_k(\epsilon)$ we obtain

$$c_k = c_k(\epsilon) + \text{cov}(T_t T_{t+k}) + (\bar{T} - \bar{T}_t)(\bar{T} - \bar{T}_{t+k}) + \frac{1}{n-k} A \quad (21)$$

In equation (21) the term $\frac{1}{n-k} A$ will in general be very small. If k is small A can be relatively large due to sampling error, but then the weight n-k is also relatively large. If n increases the conditions given by (2) will ensure that A remains relatively small. This equation shows clearly the effects of having to drop a few observations at the beginning of one series and the end of the other in order to compute the sample autocovariances. As we drop observations the mean of the series T_t $t=1,2,\dots,n-k$, given by \bar{T}_t , will differ from the mean of the series T_t $t=k+1, k+2,\dots,n$, given by \bar{T}_{t+k} , and they will both differ from the overall mean of the series as given by \bar{T} . The term $(\bar{T} - \bar{T}_t)(\bar{T} - \bar{T}_{t+k})$ will be negative and its value will increase as k increases. On the other hand $\text{cov}(T_t T_{t+k})$ will be positive and its value will decrease as k increases since the overlap between the two series will be smaller. The combination of all these effects will produce the slowly declining pattern that is observed in practice.

3 CASE OF A LINEAR TREND.

It is interesting to examine the implications of equation (21) in the case we specialise to a linear time trend

$$T_t = bt \quad (22)$$

It is found after some algebra that

$$c_k = c_k(\epsilon) + \frac{b^2}{12}(n-k+1)(n-k-1) - \frac{b^2 k^2}{4} + \frac{1}{n-k} A \quad (23)$$

We now explore the value of k for which

$$\frac{b^2}{12}(n-k+1)(n-k-1) - \frac{b^2 k^2}{4} = 0 \quad (24)$$

Discarding negative values of k the solution is

$$K^* = \frac{1}{2} \left[(3n^2 - 2)^{\frac{1}{2}} - n \right] \quad (25)$$

which for n sufficiently large can be approximated by

$$K^* = 0.366 n \quad (26)$$

This is a most surprising result. We find that the value of b , which measures the strength of the trend does not appear in the equation. It follows that the value of K remains unchanged if b changes, and that (26) is a generally valid result that applies to both deterministic and stochastic trends, whether they are linear or not.

EXAMPLE.

Figure 2 shows the logarithms of the monthly unadjusted retail price index for the period January 1974 to November 1978. It can be seen that it can be approximated reasonably well by means of a time trend. In Figure 1 the autocorrelation function for this series has been plotted. The point at which the autocorrelation is zero can be seen to be between $k = 21$ and $k = 22$. Equation (26) predicts a value of 21.6 for 59 observations.

4. REFERENCES.

/1/ O.D. ANDERSON : "On the distributional properties of serial correlations for general time processes". The Statistician, Vol. 26, Nº 3, pp. 221-230. (1977).
 /2/ O.D. ANDERSON AND J.G. DE GOOIJIER: "Distinguishing between IMA(1,1) and ARMA(1,1) models: a large scale simulation study of two particular Box-Jenkins time processes, in O.D. Anderson ed. Time Series". North Holland. (1980).

/3/ CENTRAL STATISTICAL OFFICE : Economic Trends . HMSO. London.(1974-1979).
 /4/ W.A. FULLER : "Introduction to statistical time series". Wiley. (1976).
 /5/ D.W. WICHERN : "The behaviour of the sample autocorrelation function for an IMA process". Biometrika, Vol. 60, Nº 2, pp. 235-239.

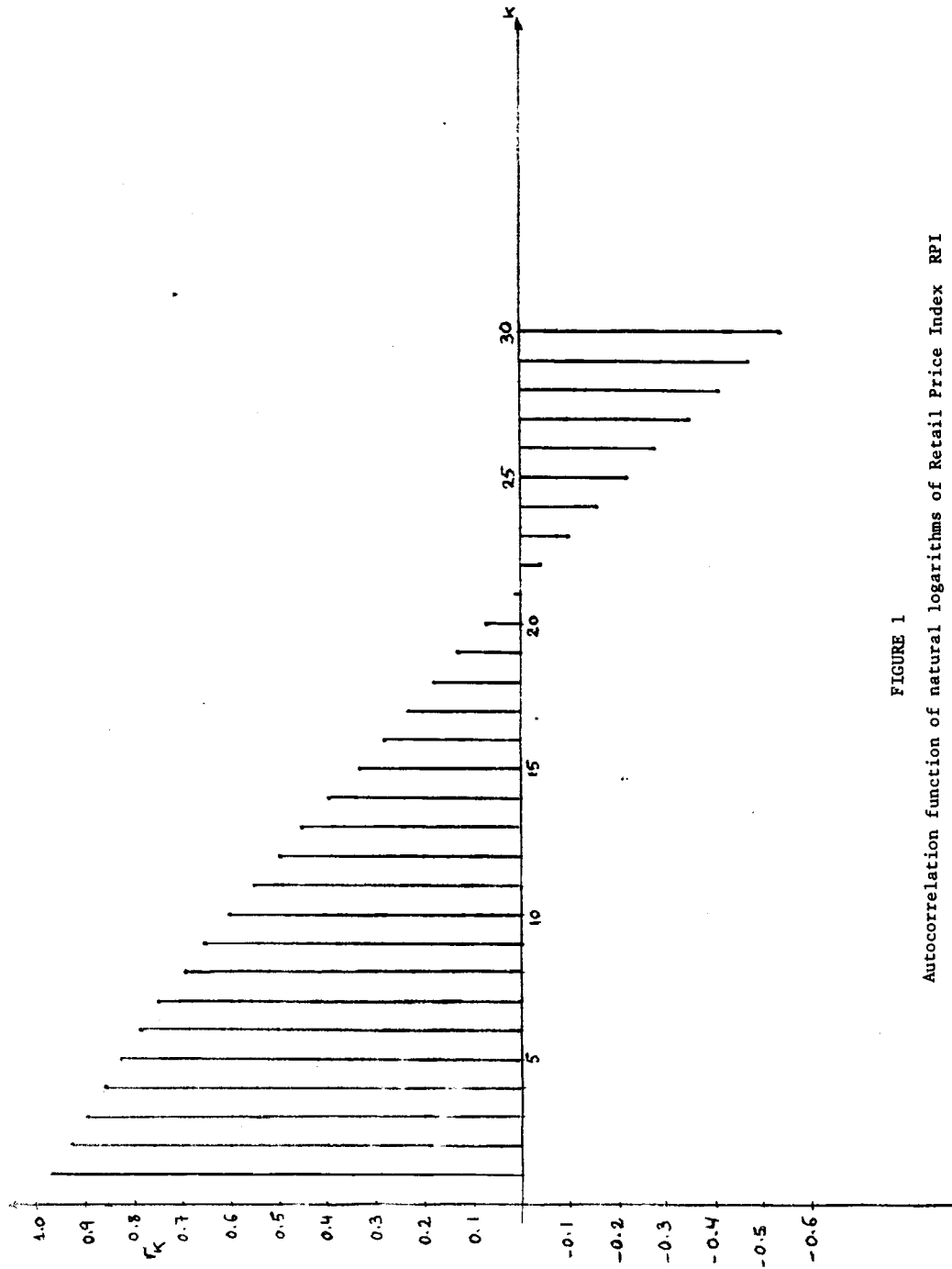


FIGURE 1
Autocorrelation function of natural logarithms of Retail Price Index RPI

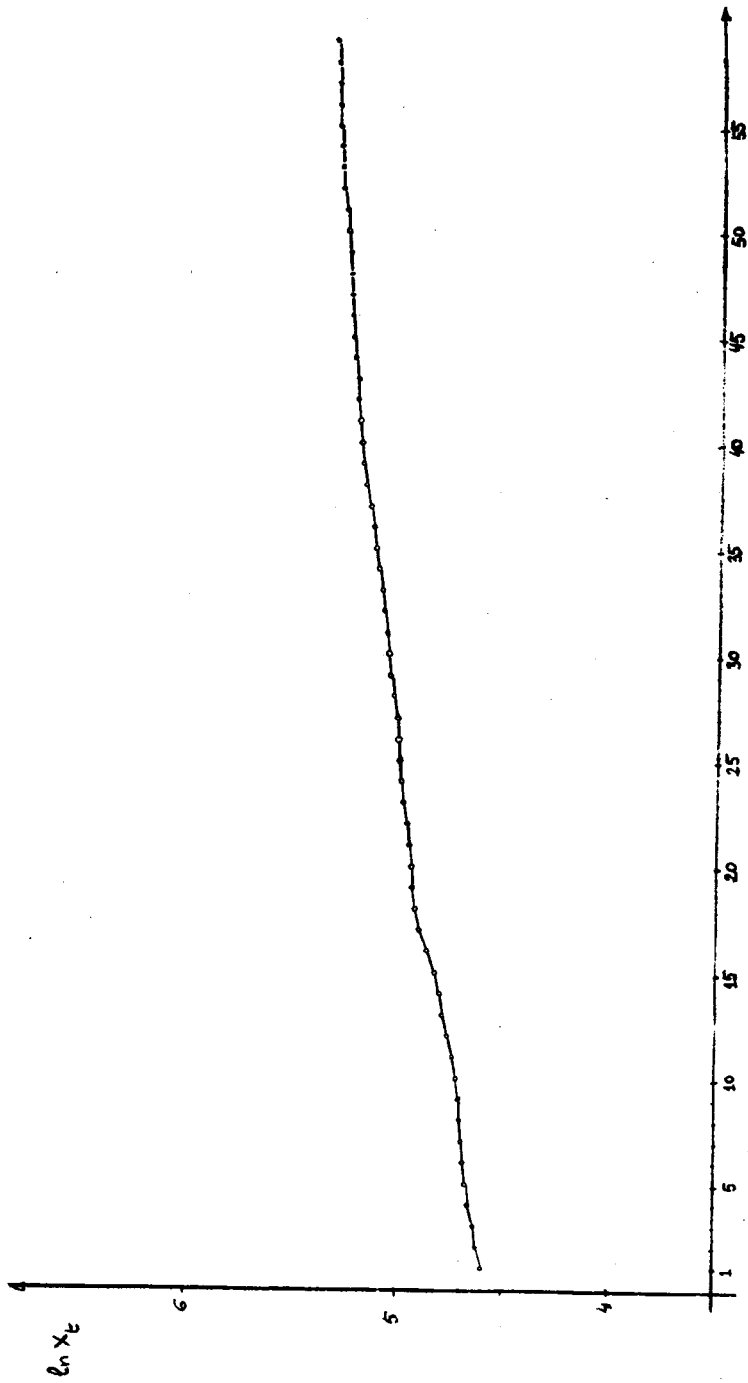


FIGURE 2

Natural logarithms of Retail Price Index RPI