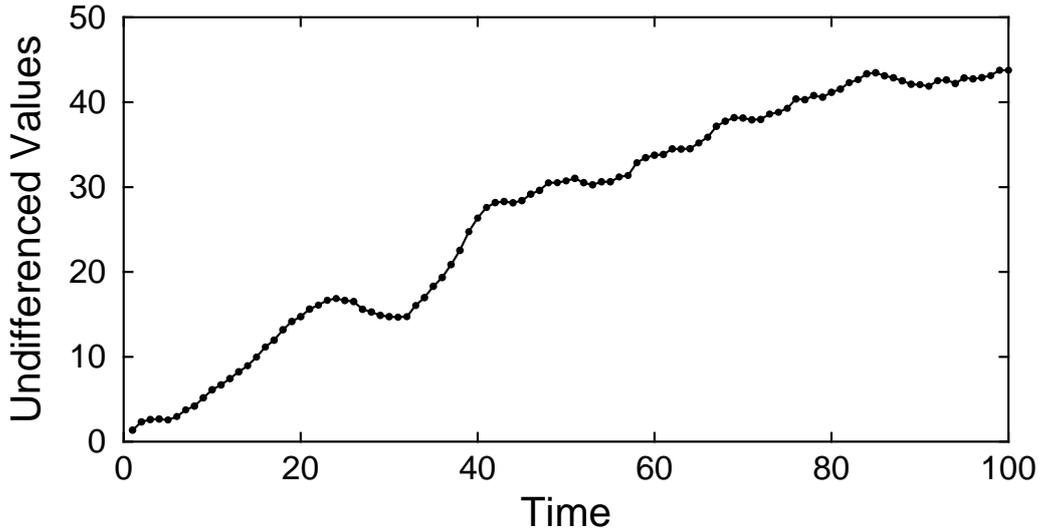




It is usual to analyze a time series for autocorrelations before proceeding to ARIMA.

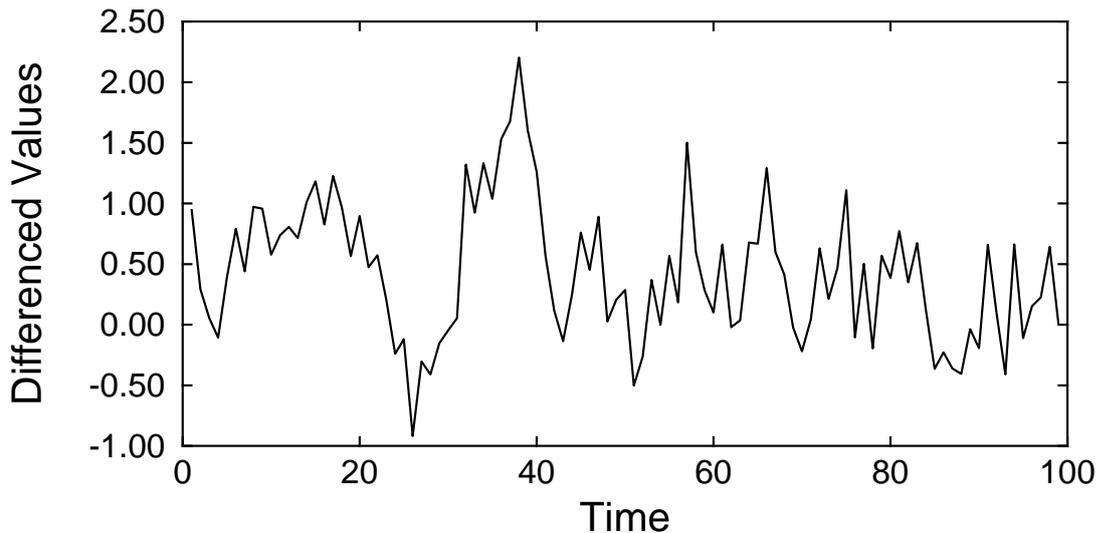
From the main SIMFIT menus choose [Statistics], [Time series] then [Lags and autocorrelations], analyze the test file times.tf1, then plot the undifferenced time series to obtain the following graph.

### Undifferenced Time Series



As there is clearly an increasing trend with these data it would be sensible to try a non-seasonal differencing of order 1 to remove a linear trend leading to the following differenced series.

### Time Series ( Non-Seasonal Differencing Order = 1)



The next table presents the results from an analysis with such a non-seasonable differencing of one and calculating for ten lags.

Current data title is: [Test file times.tf1: time series data \(J06SBF\)](#)  
 Original dimension (NX) 100  
 After differencing (NXD) 99  
 Non-seasonal order (ND) 1  
 Seasonal order (NDS) 0  
 Seasonality (NS) 0  
 Number of lags (NK) 10  
 Number of PACF (NVL) 10  
 X-mean (differenced) 0.42833  
 X-variance (differenced) 0.31521  
 Statistic (S) 83.1349  
 $P(\chi^2 \geq S)$  0.0000 *Reject  $H_0$  at 1% significance level*

Lag	R	PACF	VR	ARP
1	0.5917	0.591740	0.649844	0.391570
2	0.5258	0.270263	0.602378	0.398778
3	0.3087	-0.129867	0.592219	0.001602
4	0.1536	-0.143970	0.579944	-0.143950
5	0.0345	-0.054313	0.578233	-0.136536
6	-0.0297	0.011048	0.578162	-0.045279
7	-0.0284	0.071092	0.575240	0.147351
8	-0.0642	-0.044918	0.574080	0.130623
9	-0.1366	-0.175865	0.556324	-0.067066
10	-0.2619	-0.249824	0.521603	-0.249824

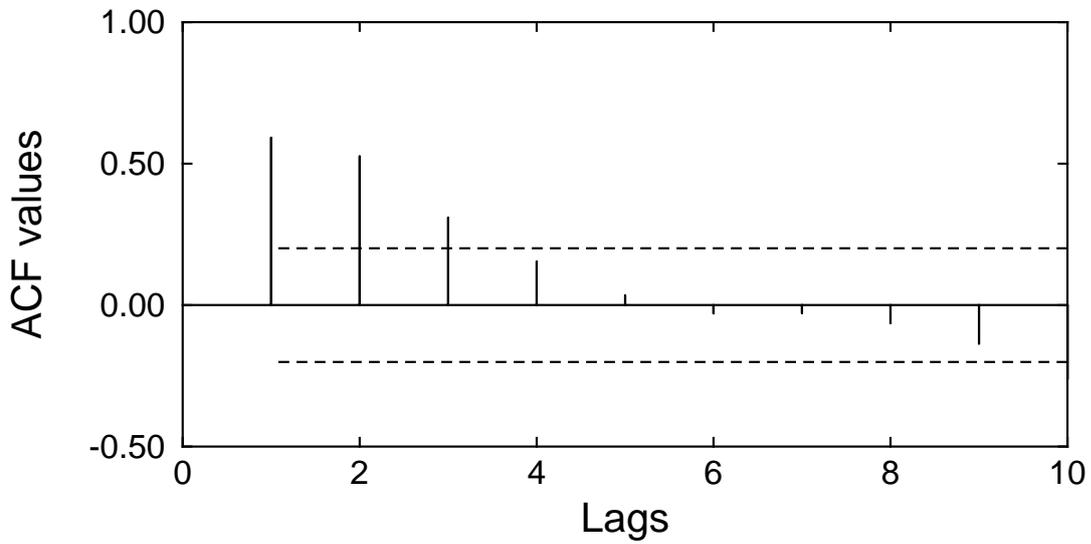
The abbreviations used in this table are defined as follows.

Abbreviation	Meaning
X	the sample (i.e. vector of data with no missing values)
NX	dimension of X
XD	transformed vector derived from X by differencing
NXD	dimension of XD (i.e. $NXD = NX - ND - NS \cdot NDS$ )
ND	order of non-seasonal differencing
NDS	order of seasonal differencing
NS	seasonality of seasonal differencing
NK	number of lags requested
NL	number of partial correlations requested
NVL	number of valid partial correlations
R	calculated autocorrelation coefficients
PACF	calculated partial autocorrelation coefficients
VR	calculated predictor error variance ratios
ARP	calculated autoregressive parameters of maximum order
S	statistic for testing $H_0$ : all autocorrelations zero
Note	if $ND = NDS = NS = 0$ , the original sample X is analyzed

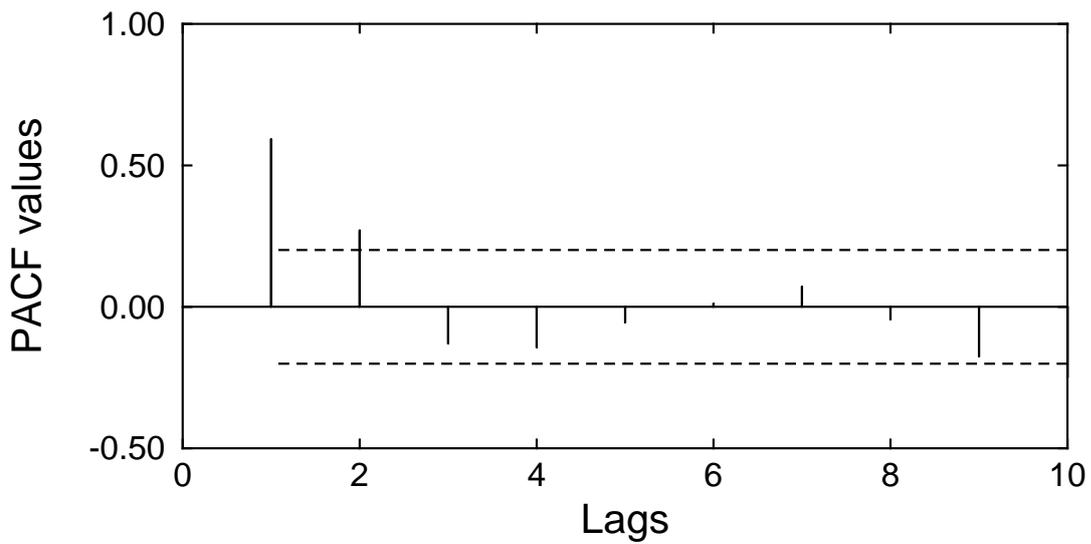
From these results the  $\chi^2$  test clearly indicates that not all correlations are zero, so it is useful to proceed to comparing the strengths of autocorrelations and partial autocorrelations as functions of the lags.

Note that, in the next figures, the approximate 95% confidence limits are estimated as  $2/\sqrt{n}$ , where  $n$  is the sample size after differencing (if any).

### Autocorrelation Function



### Partial Autocorrelation Function



These figures illustrate what is summarized in the previous table, that only the correlations with lags of one and two are highly significant.

## Theory

A time series is a vector  $x(t)$  of  $n > 1$  observations  $x_i$  obtained at a sequence of points  $t_i$ , e.g., times, distances, etc., at fixed intervals  $\Delta$ , i.e.

$$\Delta = t_{i+1} - t_i, \text{ for } i = 1, 2, \dots, n-1,$$

and it is assumed that there is some seasonal variation, or other type of autocorrelation to be estimated.

A linear trend can be removed by first order differencing

$$\nabla x_t = x_t - x_{t-1},$$

while seasonal patterns of seasonality  $s$  can be eliminated by first order seasonal differencing

$$\nabla_s x_t = x_t - x_{t-s}.$$

Note that differencing of orders  $d = ND$ ,  $D = NDS$ , and seasonality  $s = NS$  may be applied repeatedly to a series so that

$$w_t = \nabla^d \nabla_s^D x_t$$

will be shorter, of length  $NXD = n - d - D \times s$ , and will extend for  $t = 1 + d + D \times s, \dots, NX$ .

Non-seasonal differencing up to order  $d$  is calculated sequentially using

$$\begin{aligned} \nabla^1 x_i &= x_{i+1} - x_i && \text{for } i = 1, 2, \dots, n-1 \\ \nabla^2 x_i &= \nabla^1 x_{i+1} - \nabla^1 x_i && \text{for } i = 1, 2, \dots, n-2 \\ \dots & && \\ \nabla^d x_i &= \nabla^{d-1} x_{i+1} - \nabla^{d-1} x_i && \text{for } i = 1, 2, \dots, n-d \end{aligned}$$

while seasonal differencing up to order  $D$  is calculated by the sequence

$$\begin{aligned} \nabla^d \nabla_s^1 x_i &= \nabla^d x_{i+s} - \nabla^d x_i && \text{for } i = 1, 2, \dots, n-d-s \\ \nabla^d \nabla_s^2 x_i &= \nabla^d \nabla_s^1 x_{i+s} - \nabla^d \nabla_s^1 x_i && \text{for } i = 1, 2, \dots, n-d-2s \\ \dots & && \\ \nabla^d \nabla_s^D x_i &= \nabla^d \nabla_s^{D-1} x_{i+s} - \nabla^d \nabla_s^{D-1} x_i && \text{for } i = 1, 2, \dots, n-d-D \times s. \end{aligned}$$

Note that, as indicated in the previous tables, either the original sample  $X$  of length  $NX$ , or a differenced series  $XD$  of length  $NXD$ , can be analyzed interactively, by simply adjusting  $ND$ ,  $NDS$ , or  $NS$ . Also the maximum number of autocorrelations  $NK < NXD$  and maximum number of partial autocorrelations  $L \leq NK$ , can be controlled, although the maximum number of valid partial autocorrelations  $NVL$  may turn out to be less than  $L$ . Now, defining either  $x = X$ , and  $n = NX$ , or else  $x = XD$  and  $n = NXD$  as appropriate, and using  $K = NK$ , the mean and variance are recorded, plus the autocorrelation function  $R$ , comprising the autocorrelation coefficients of lag  $k$  according to

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

If  $n$  is large and much larger than  $K$ , then the  $S$  statistic

$$S = n \sum_{k=1}^K r_k^2$$

has a chi-square distribution with  $K$  degrees of freedom under the hypothesis of zero autocorrelation, and so it can be used to test that all correlations are zero.

The partial autocorrelation function *PACF* has coefficients at lag  $k$  corresponding to  $p_{k,k}$  in the autoregression

$$x_t = c_k + p_{k,1}x_{t-1} + p_{k,2}x_{t-2} + \cdots + p_{k,k}x_{t-k} + e_{k,t}$$

where  $e_{k,t}$  is the predictor error, and the  $p_{k,k}$  estimate the correlation between  $x_t$  and  $x_{t+k}$  conditional upon the intermediate values  $x_{t+1}, x_{t+2}, \dots, x_{t+k-1}$ . Note that the parameters change as  $k$  increases, and so  $k = 1$  is used for  $p_{1,1}$ ,  $k = 2$  is used for  $p_{2,2}$ , and so on.

These parameters are determined from the Yule-Walker equations

$$r_i = p_{k,1}r_{i-1} + p_{k,2}r_{i-2} + \cdots + p_{k,k}r_{i-k}, \quad i = 1, 2, \dots, k$$

where  $r_j = r_{|j|}$  when  $j < 0$ , and  $r_0 = 1$ .

An iterative technique is used and it may not always be possible to solve for all the partial autocorrelations requested. This is because the predictor error variance ratios *VR* are defined as

$$\begin{aligned} v_k &= \text{Var}(e_{k,t}) / \text{Var}(x_t) \\ &= 1 - p_{k,1}r_1 - p_{k,2}r_2 - \cdots - p_{k,k}r_k, \end{aligned}$$

unless  $|p_{k,k}| \geq 1$  is encountered at some  $k = L_0$ , when the iteration terminates, with  $NVL = L_0 - 1$ .

The Autoregressive parameters of maximum order *ARP* are the final parameters  $p_{L,j}$  for  $j = 1, 2, \dots, NVL$  where  $NVL$  is the number of valid partial autocorrelation values, and  $L$  is the maximum number of partial autocorrelation coefficients requested, or else  $L = L_0 - 1$  as before in the event of premature termination of the algorithm.