Contract no: 20851A2/1/020

Time Series: General

David R. Brillinger

University of California, Berkeley, California, USA

Time series: a stretch of values on the same scale indexed by a time-like parameter. The basic data and parameters are functions.

Time series take on a dazzling variety of shapes and forms, indeed there are as many time series as there are functions of real numbers. Some common examples of time series forms are provided in Figure 1. One notes periods, trends, wandering and integer-values. The time series such as those in the Figure may be contemporaneous and a goal may be to understand the interrelationships.

Concepts and fields related to time series include: longitudinal data, growth curves, repeated measures, econometric models, multivariate analysis, signal processing and systems analysis.

The field, *time series analysis*, consists of the techniques which when applied to time series lead to improved knowledge. The purposes include summary, decision, description, prediction.

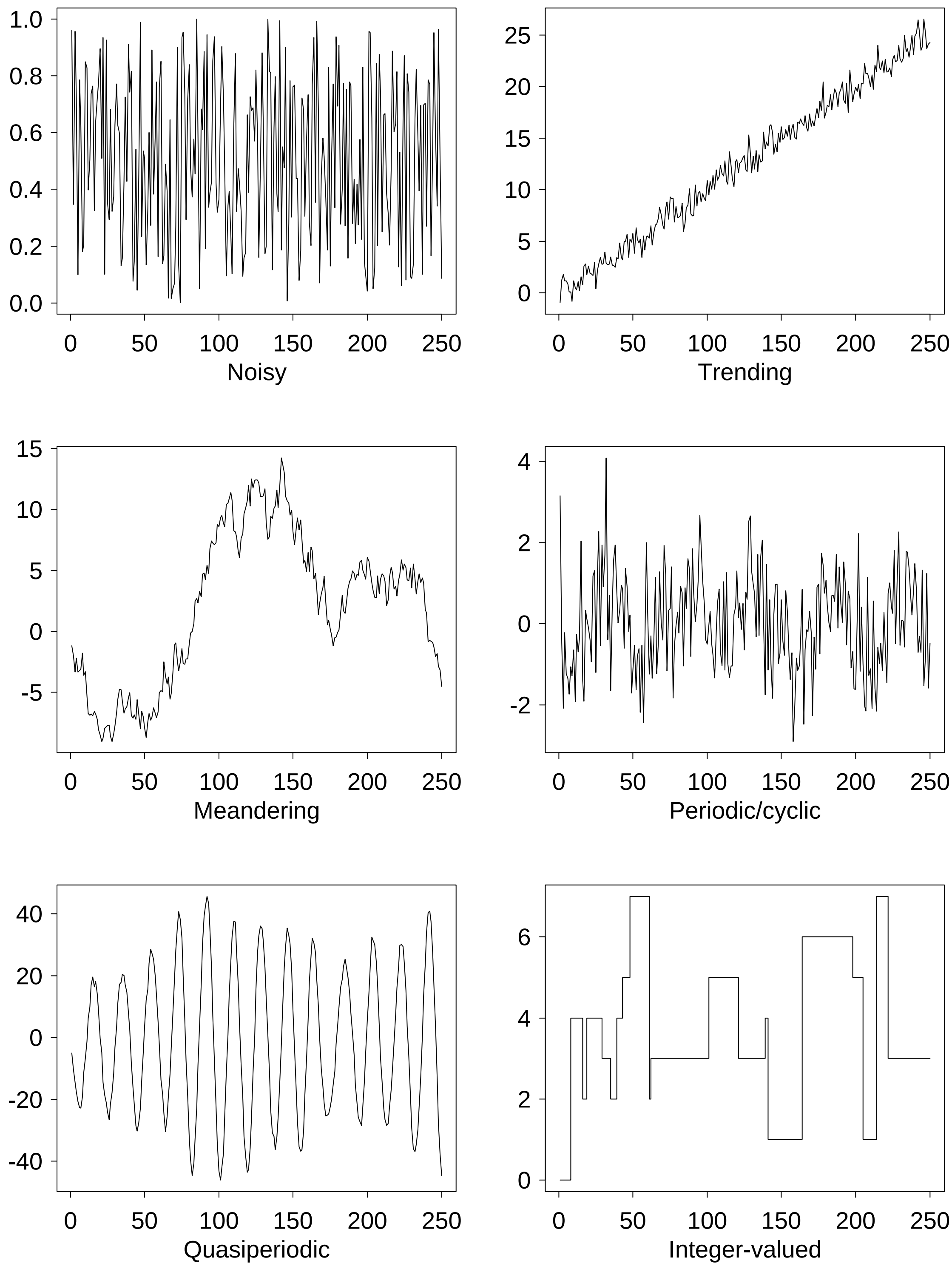

Figure 1

Some different types of time series.

The field has a theoretical side and an applied side. The former is part of the theory of stochastic processes (e.g. representations, prediction, information, limit theorems) while applications often involve extensions of techniques of "ordinary" statistics e.g. regression, analysis of variance, multivariate analysis, sampling. The field is renowned for jargon and acronyms - white noise, cepstrum, ARMA, ARCH, see Granger [16].

1 Importance

"... but time and chance happeneth to them all." *Ecclesiastes*

Time series ideas appear basic to virtually all activities. Time series are used by Nature and humans alike for communication, description and visualisation. Because *time* is a physical concept, parameters and other characteristics of mathematical models for time series can have real-world interpretations. This is of great assistance in the analysis and synthesis of time series.

Time series are basic to scientific investigations. There are: circadian rhythms, seasonal behaviors, trends, changes and evolving behavior to be studied and understood. Basic questions of scientific concern are formulated in terms of time series concepts - Predicted value? Leading? Lagging? Causal connection? Description? Association? Autocorrelation? Signal? Seasonal effect? New phenomenon? Control? Periodic? Changing? Trending? Hidden period? Cycles?

Because of the tremendous variety of possibilities, substantial simplifications are needed in many time series analyses. These may include assumptions of stationarity, mixing or asymptotic independence, normality, linearity. Luckily such assumptions often appear plausible in practice.

The subject of time series would be important if for no other reason than that it provides means of examining the basic assumption of statistical independence invariably made in ordinary statistics. One of the first commonly used procedures for this problem was the Durbin-Watson test [11]. The autocovariance and spectrum functions, see below, are now often used in this context.

2 History

Contemporary time series analysis has substantial beginnings in both the physical and social sciences. Basic concepts have appeared in each subject and made their way to the other with consequent transfer of technology. Historical researchers important in the development of the field include: Thiele [29], Hooker [26], Einstein [12], Wiener [41], Yule [44], Fisher [15], Tukey [38], Whittle [40], Bartlett [2]. Books particularly influential in the social sciences include Moore [31] and Davis [8]. Nowadays many early analyses appear naive. For example, Beveridge in 1920 listed some 19 periods for a wheat price index running from 1500 to 1869 [3]. It is hard to imagine the presence of so much regular behavior in such a series. When statistical uncertainty is estimated using later day techniques most of these periods appear insignificant,

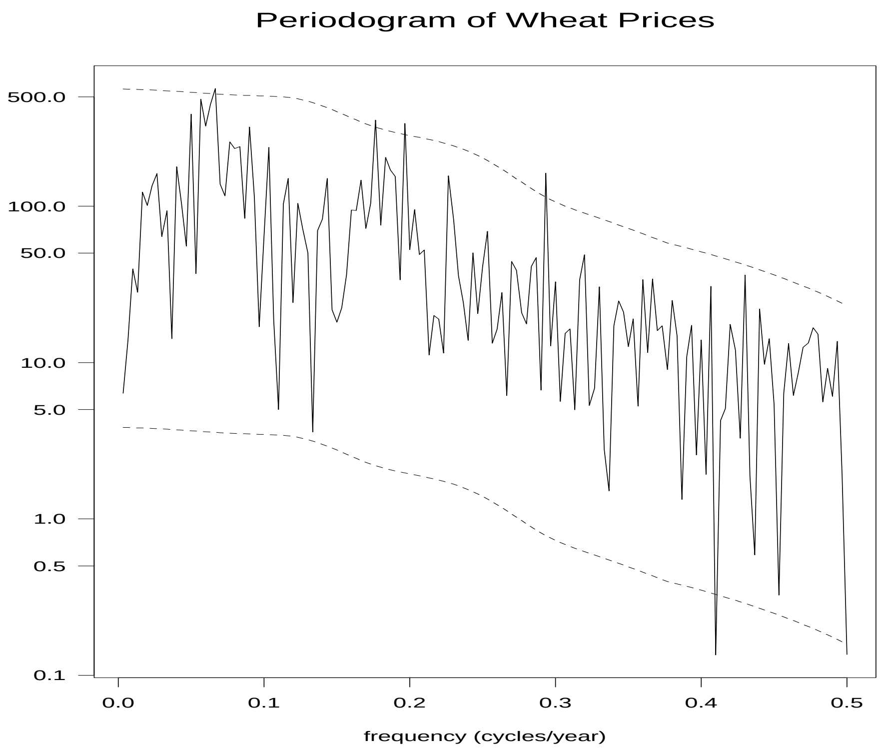

Figure 2

The dashed lines provide 95 percent bounds about a central curve.

see the periodogram with 95 percent error bounds in Figure 2. Many historical references are included in Wold [43].

Historians of science have made some surprising discoveries concerning early work with time series. An example is presented in Tufte [37]. He shows a time series plot from the 10th or 11th century AD. This graph is speculated to provide movements of the planets and sun. It is remarkable for its cartesian type character. More generally Tufte remarks, following a study of newspapers and magazines, that "The time-series plot is the most frequently used form of graphic design." Casual observation certainly supports Tufte's study.

Important problems that were addressed in the twentieth century include: seasonal adjustment, leading and laging indicators and index numbers. Paradigms that were developed include:

$$series \;=\; signal \;+\; noise$$

$$series \;=\; trend \;+\; seasonal \;+\; noise$$

$$series \;=\; sum\ of\ cosines \;+\; noise$$

$$series \;=\; regression\ function \;+\; noise$$

These conceptualizations have been used for forecasting, seasonal adjustment and description amongst other things. There are surprises, e.g. ordinary least squares is surprisingly efficient in the time series case, see Grenander and Rosenblatt [19]. Other books from the 1950s and 1960s that proved important are Blackman and Tukey [4], Granger and Hatanaka [18], Box and Jenkins [5]. Important papers from the era include: Akaike [1], Hannan [22], Parzen [33].

An important development that began in the 1950s and continues today is the use of recursive computation in the state space model, see Kalman [27], Harvey [23], Shumway and Stoffer [35] and Durbin and Koopman [10].

3 Basics

3.1 Concepts

There are a number of concepts that recur in time series work. Already defined is the *time series*, a stretch of values on the same scale indexed by a time parameter. The time parameter may range over the positive and negative integers or all real numbers or subsets of these. *Time series data* refers to a segment of a time series. A time series is said to have a *trend* when there is a slowly evolving change. It has a *seasonal* component when some cyclic movement of period one year is present. (The *period* of a cyclic phenomenon is the amount of time for it to repeat itself. Its *frequency* is the reciprocal of the period.)

There is an algebra of mathematical operations that either Nature or an analyst may apply to a time series $\{X(t)\}$ to produce another series $\{Y(t)\}$. Foremost is the *filter*, or *linear time invariant operation* or *system*. In the case of discrete equi-spaced time this may be represented as

$$Y(t) \ = \ \sum_u a(u)X(t-u)$$

$t, u \ = \ 0, \pm 1, ...$ An example is the running mean used to smooth a series. The functions $X(.), \ Y(.)$ may be vector-valued and $a(.)$ may be matrix-valued. In the vector-valued case feedback may be present. The sequence $\{a(u)\}$ is called the *impulse response function*. The operation has the surprising property of taking a series of period P into a series of the same period P. The filter is called *realizable* when $a(u) = 0$ for $u < 0$. Such filters appear in causal systems and when the goal is prediction.

The above ideas concern both deterministic and random series. The latter prove to be important in developing solutions to important problems. Specifically it often proves useful to view the subject of time series as part of the general theory of stochastic processes, that is indexed families of random variables. One writes $\{Y(t,\omega), t \; in \; V\}$, with ω a random variable and V a set of times. Time series data are then viewed as a segment, $\{Y(t,\omega_0), t = 0, ..., T-1\}$ of the realization labelled by ω_0, the obtained value of ω. Stochastic time series $\{Y(t), t = 0, \pm 1, \pm 2, ...\}$ are sometimes conveniently described by finite dimensional distributions such as

$$Prob\{Y(t) \leq y\} \; and \; Prob\{Y(t_1) \leq y_1, \; Y(t_2) \leq y_2\}$$

This is the case for time series values with joint normal distributions.

Time series may also be usefully described or generated by stochastic models involving the independent identically distributed random variables of ordinary statistics.

Stochastic models may be distinguished as parametric or nonparametric. Basic parameters of the nonparametric approach include: moments, joint probability and density functions, mean functions, autocovariance and crosscovariance functions, power spectra.

One basic assumption of time series analysis is that of *stationarity*. Here the choice of time origin does not affect the statistical properties of the process. For example the mean level of a stationary series is constant. Basic to time series analysis is handling temporal dependence. To this end one can define the crosscovariance function of the series X and Y at lag u as the covariance of the values $X(t+u)$ and $Y(t)$. In the stationary case this function does not depend on t. In an early paper, [26], Hooker computed an estimate of this quantity. Another useful parameter is the power spectrum, a display of the intensity or variability of the phenomenon versus period or frequency. It may be defined as the Fourier transform of the autocovariance function. The power spectrum proves useful in displaying the serial dependence present, in discovering periodic phenomena and in diagnosing possible models for a series.

In the parametric case there are the autoregressive moving average (ARMA) series. These are a regression-type models expressing the value $Y(t)$ as a linear function of a finite number of past values $Y(t-1), Y(t-2), ...$ of the series and the values $\epsilon(t), \epsilon(t-1), ...$ of a sequence of independent identically distributed random variables. ARMAs have proved particularly useful in problems of forecasting.

Contemporary models for time series are often based upon the idea of *state*. This may be defined as the smallest entity summarizing the past history of the process. There are two parts to the state space model. First, a state equation that describes how the value of the state evolves in time. This equation may contain a purely random element. Second there is an equation indicating how the available measurements at that time t come from the state value at that time. It too may involve a purely random element. The concept of the state of a system has been basic to physics for many years in classical dynamics and quantum mechanics. The idea was seized upon by control engineers, e.g. Kalman [27] in the late fifties. The econometricians realized its usefulness in the early eighties, see Harvey [23].

A number of specific probability models have been studied in depth, including the Gaussian, the ARMA, the bilinear [17], various other nonlinear [36], long and short memory, ARMAX, ARCH [13], hidden Markov [30], random walk, stochastic differential equations [20,34] and the periodically stationary.

A list of journals where these processes are often discussed is included at the end of this entry.

3.2 Problems

There are scientific problems and there are associated statistical problems that arise. Methods have been devised for handling these. The scientific problems include: smoothing, prediction, association, index numbers, feedback and control. Specific statistical problems arise directly. Among these are including explanatories in a model, estimation of parameters such as hidden frequencies, uncertainty computation, goodness of fit and testing.

Special difficulties arise. These include: missing values, censoring, measurement error, irregular sampling, feedback, outliers, shocks, signal-generated noise, trading days, festivals, changing seasonal pattern, measurement error, aliasing, data observed in two series at different time points.

Particularly important are the problems of association and prediction. The former asks the questions of whether two series are somehow related and what is the strength of any association? Measures of association include: the cross-corellation and the coherence functions. The prediction problem concerns the forecasting of future values. There are useful mathematical formulations of this problem but because of unpredictable human intervention there are situations where guesswork seems just as good.

Theoretical tools employed to address the problems of time series analysis include: mathematical models, asymptotic methods, functional analysis and transforms.

4 Methods

4.1 Descriptive

Descriptive methods are particularly useful for exploratory and summary purposes. They involve graphs and other displays, simple statistics such as means, medians and percentiles and the techniques of exploratory data analysis, [39].

The most common method of describing a time series is by presenting a graph, see Figure 1. Such graphs are basic for communication and assessing a situation. There are different types. Cleveland, [7], mentions the connected, symbol, connected-symbol and vertical-line displays in particular. Figure 1 presents connected graphs.

Descriptive values derived from time series data include extremes, turning points, level crossings and the periodogram. See Figure 2 for an example of this last. Descriptive methods typically involve generating displays using manipulations of time series data via operations such as differencing, smoothing and narrow band filtering.

A display with a long history, [28,43], is the Buys-Ballot table. Among other things it is useful for studying the presence and character of a phenomenon of period P such as a circadian rhythm. One creates a matrix with entry $Y((i-1)P+j)$ in row i, column $j = 1,...,P$ and then, for example, one computes column averages. These values provide an estimate of the effect of period P. The graphs of the individual rows may be stacked beneath each other in a display. This is useful for discerning slowly evolving behaviour.

Descriptive methods may be time-side (as when a running mean is computed), frequency-side (as when a periodogram is computed) or hybrid (as when a sliding window periodogram analysis is employed).

4.2 Parameter estimation

The way to a solution of many time series problems is via the setting down of a stochastic model. Parameters are constants of unknown values included in the model. They are important because substantial advantages arise when one works within the context of a model. These advantages include: estimated standard errors, efficiency and communication of results. Often parameter estimates are important because they are fed into later procedures, e.g. formulas for forecasting.

General methods of estimating parameters include: method of moments, least squares and maximum likelihood. An important time series case is that of harmonic regression. It was developed in Fisher [15] and Whittle [40].

There are parametric and nonparametric approaches to estimation. The parametric has the advantage that if the model is correct, then the estimated coefficients have considerably smaller standard errors.

4.3 Uncertainty estimation

Estimates without some indication of their uncertainty are not particularly useful in practice. There are a variety of methods used in time series analysis to develop uncertainty measures. If maximum likelihood estimation has been used there are classic (asymptotic) formulas. The delta-method or method of linearization is useful if the quantity of interest can be recognized to be a smooth functon of other quantities whose variability can be estimated directly. Methods of splitting the data into segments, such as the jackknife, can have time series variants. A method currently undergoing enjoying considerable investigation is the bootstrap, [9]. The assumptions made in validating these methods are typically that the series involved is stationary and mixing.

4.4 Seasonal adjustment

Seasonal adjustment may be defined as the attempted removal of obscuring unobservable annual components. There are many methods of seasonal adjustment, [32], including state space approaches [23].

The power spectrum provides one means of assessing the effects of various suggested adjustment procedures.

4.5 System identification

System identification refers to the problem of obtaining a description or model of a system on the basis of a stretch of input to and the corresponding output from a system of interest. The system may be assumed to be linear time invariant as defined above.

In designed experiments the input may be a series of pulses, a sinusoid or white noise. In a natural experiment the input is not under the control of the investigator and this leads to complications in the interpretation of the results.

System identification relates to the issue of causality. In some systems one can turn the input off and on and things are clearer.

4.6 Computing

Important computer packages for time series analysis include: Splus, Matlab, Mathematica, SAS, SPSS, TSP, STAMP. Some surprising algorithms have been found: the fast Fourier transform (FFT), fast algorithms from computational geometry, Monte Carlo methods and the Kalman-Bucy filter. Amazingly a variant of this last was employed by Thiele in 1880, [29], while the FFT was known to Gauss in the early 1800s, [25].

5 Current theory and research

Much of what is being developed in current theory and research is driven by what goes on in practice in time series analyses. What is involved in a time series analysis? The elements include: the question, the experiment, the data, plotting the data, the model, model validation and model use. The importance of recognizing and assessing the basic assumptions is fundamental.

The approach adopted in practice often depends on how much data are available. In the case that there are a lot of data procedures that are in some sense inefficient, are often used effectively. A change from the past is that contemporary analysis often results from the appearance of very large fine data sets. The amount of data can seem limitless as for example in the case of records of computer tasks. There are many hot research topics. One can mention: exploratory data analysis techniques for very large data sets, so-called improved estimates, testing (association? cycle present?), goodness of fit diagnostics. There are the classical and Bayesian approaches, the parametric, semi-parametric and nonparametric analyses, the problem of dimension estimation and that of re-expression of variables.

Current efforts include research into: bootstrap variants, long-memory processes, long-tailed distributions, nonGaussian processes, unit roots, nonlinearities, better approximations to distributions, demonstrations of efficiency, self-similar processes, scaling relationships, irregularly observed data, doubly stochastic processes as in hidden Markov, cointegration, disaggregation, cyclic stationarity, wavelets and particularly inference for stochastic differential equations.

Today's time series data values may be general linear model type [14], e.g. ordinal, state-valued, counts, proportions, angles, ranks. They may be vectors. They may be functions. The time label t may be location in space or even a function. The series may be vector-valued. The data may have been collected in an experimental design.

There are some surprises: the necessity of tapering and prewhitening to reduce bias, the occurrence of aliasing, the high efficiency of ordinary least squares estimates in the correlated case and the approximate independence of empirical

Fourier transform values at different frequencies [6].

6 Future directions

It seems clear that time series research will continue on all the existing topics as the assumptions on which any existing problem solution has been based appear inadequate. Further it can be anticipated that more researchers from nontraditional areas will become interested in the area of time series as they realize that the data they have collected, or will be collecting, are correlated in time.

Researchers can be expected to be even more concerned concerned with the topics of nonlinearity, conditional heteroscedasticity, inverse problems, long memory, long tails, uncertainty estimation, inclusion of explanatories, new analytic models and properties of the estimates when the model is not true. The motivation for the last is that time series with unusual structure seem to appear steadily. An example is the data being collected automatically via the World Wide Web. Researchers can be anticipated to be seeking extensions of existing time series methods to processes with more general definitions of the time label - spatial, spatial-temporal, functional, angular. At the same time they will work on processes whose values are more general, even abstract.

More efficient, more robust and more applicable solutions will be found for existing problems. Techniques will be developed for dealing with special difficultes such as missing data, nonstationarity, outliers. Better approximations to the distributions of time series based statistics will be developed.

Many have stressed the advantages of linear system identification via white noise input. Wiener [42] stressed the benefits of using Gaussian white noise input. This idea has not been fully developed. Indeed data sets obtained with this input will continually yield to novel analytic methods as they make their appearance.

The principal journals in which newly developed statistical methods for time series are presented and studied include: *J. Time Series Analysis*, *Annals of Statistics*, *Stochastic Processes and Their Applications*, *J. American Statistical Association*, *Econometrica*, *IEEE Trans. Signal Processing*.

Bibliography

[1] Akaike, H 1965 On the statistical estimation of the frequency response function of a system having multiple input. *Ann. Inst. Stat. Mth.* **17** 185-210

[2] Bartlett, M S 1966 *Stochastic Processes* Cambridge University Press, Cambridge

[3] Beveridge W H 1922 Wheat prices and rainfall in western Europe. *J. Royal Statistical Soc* **85** 412-459

[4] Blackman, R and Tukey, J W 1959 *The Measurement of Power Spectra* Dover, New York.

[5] Box, G E P and Jenkins, G M 1970. *Time Series Analysis; Forecasting and Control* Holden-Day, San Francisco.

[6] Brillinger, D R 1975 *Time Series: Data Analysis and Theory* Holden-Day, San Francisco

[7] Cleveland W S 1985 *The Elements of Graphing Data.* Wadsworth, Pacific Grove

[8] Davis H T 1941 *The Analysis of Economic Time Series.* Principia, Bloomington

[9] Davison, A and Hinkley, D 1997 *Bootstrap Methods and Their Application* Cambridge University Press, Cambridge

[10] Durbin, J. and Koopman, S. J. 2000 *Time Series Analysis by State Space Methods* Oxford University Press, Oxford

[11] Durbin, J and Watson, G S 1950,1 Testing for Serial correlation in least squares regression I,II *Biometrika* **37,38** 409-428, 159-177

[12] Einstein, A. 1914 Méthode pour la détermination des valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières. *Arch. Sci. Phys. et Naturelles Ser. 4* **37** 254-255.

[13] Engle, R 1995 *ARCH Selected Readings* Oxford University Press, Oxford

[14] Fahrmeir, L and Tutz, G 1994 *Multivariate Statistical Modelling Based on Generalized Linear Models* Springer, New York

[15] Fisher R A 1929 Tests of significance in harmonic analysis *Proc. Royal Soc. London A* **125** 54-59

[16] Granger, C W J 1982 Acronyms in time series analysis *J. Time Series Analysis* **2** 103-108

[17] Granger, C W J 1978 *An Introduction to Bilinear Time Series Models* Vandenhoeck and Ruprecht, Gottingen

[18] Granger, C W J and Hatanaka, M 1964 *Spectral Analysis of Economic Time Series* Princeton University Press, Princeton

[19] Grenander, U and Rosenblatt, M 1957 *Statistical Analysis of Stationary Time Series* Wiley, New York

[20] Guttorp, P 1995 *Stochastic Modelling of Scientific Data* Chapman and Hall, London

[21] Hannan, E J 1960 *Time Series Analysis* Methuen, London

[22] Hannan, E J 1963 Regression for time series *Time Series Analysis* Ed. M. Rosenblatt Wiley, New York

[23] Harvey A C 1989 *Forecasting, Structural Time Series Models and the Kalman Filter* Cambridge University Press, Cambridge

[24] Harvey, A C, Koopman, S J and Penzer, J 1998 Messy time series: a unified approach. *Advances in Econometrics* **13** Eds. T. B. Fomby and R. C. Hill JAI Press, New York.

[25] Heideman, M T, Jondon, D H and Burrus, C S 1984 Gauss and the history of the fast Fourier transform *IEEE Acoust. Speech Signal Proc. Magazine* **1** 14-21

[26] Hooker R H 1901 Correlation of the marriage-rate with trade *J Royal Statist Soc* **64** 696-703

[27] Kalman, R E 1963 Mathematical description of linear dynamical systems *SIAM J. Control* **1**, 152-192

[28] Laplace P S 1825 *Traité de Méchanique Céleste* **5** Bachelier, Paris

[29] Lauritzen, S L 1981 Time series analysis in 1880: a discussion of contributions made by T. N. Thiele *Internat. Statist. Review* **49** 319-331

[30] MacDonald, I and Zucchini, W 1997 *Hidden Markov and Other Models for Discrete-valued Time Series* Chapman and Hall, London.

[31] Moore, H L 1914. *Economic Cycles: Their Law and Cause.* Macmillan, New York.

[32] National Bureau of Economic Research/Bureau of the Census 1976 *Conference on Seasonal Analysis of Economic Time Series* Dept. of Commerce, Bureau of the Census, Washington, DC

[33] Parzen, E 1961 An approach to time series *Ann. Math. Statist.* **32** 951-989

[34] Prakasa Rao, B. L. S. 1999 *Statistical Inference for Diffusion Type Processes* Arnold, London

[35] Shumway, R. H. and Stoffer, D. S. 2000 *Time Series Analysis and Its Applications* Springer, New York

[36] Tong, H 1990 *Non-linear Time Series : a Dynamical System Approach* Oxford University Press, New York

[37] Tufte, E R 1983 *The Elements of Graphing Data* Graphics Press, Cheshire, Connecticut

[38] Tukey, J W 1984 *The Collected Works of John W. Tukey I-II Time Series.* Wadsworth, Monterey

[39] Tukey, J W 1977 *Exploratory Data Analysis* Addision-Wesley, Reading

[40] Whittle, P 1951 *Hypothesis Testing and Time Series Analysis* Almqvist and Wiksell, Uppsala

[41] Wiener N 1949 *Time Series* MIT Press, Cambridge

[42] Wiener, N 1958 *Nonlinear Problms in Random Theory* MIT Press, Cambridge

[43] Wold, H O A *Bibliography on Time Series and Stochastic Processes* MIT Press, Cambridge

[44] Yule, G U 1927 On the method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Phil. Trans. A* **226** 267-298