STATS 3005

TIME SERIES III

Lecturer: Associate Professor Patty Solomon
School of Mathematical Sciences
The University of Adelaide

Semester 2, 2004

These outline notes are copies of the overhead transparencies shown in lectures, and are intended as a guide to this course.

1 INTRODUCTION

1.1 Some history and examples

The time series plot is the most frequently used form of graphic design.

- The oldest known attempt to show changing values graphically dates from the 10th century and illustrates the inclinations of the planetary orbits as a function of time. Figure 1.1 is from a manuscript of Macrobius, commentary of Cicero's *In Somnium Scipionis*, 10th century.

- It was not until the late 1700's that time series charts began to appear in scientific writing. The two great inventors of modern graphical designs were Lambert (1728–1777) a Swiss-German scientist and mathematician, and William Playfair (1759–1823) an English political economist.

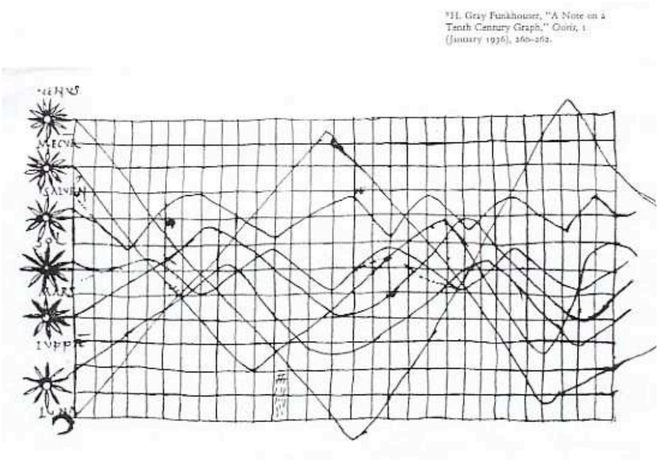

Fig. 1.1: Inclinations of the planetary orbits as a function of time. From E.R. Tufte (1983) *The visual display of quantitative information.*

- The first known time series using economic data was published in Playfair's book *The commerical and political atlas*, London 1786.

  In later work, Playfair (1821) addressed the question of whether the price of wheat had increased relative to wages. He plotted three parallel time series: prices, wages, and the reigns of British kings and queens; see Figure 1.2.

- Up to 1925, a time series was regarded as generated deterministically. But R.A. Fisher made clear that measured assessments of variability are at the heart of quantitative reasoning (*Statistical Methods for Research Workers*, Edinburgh 1925).

- In 1927, Udney Yule broke new ground in the analysis of sunspot data.

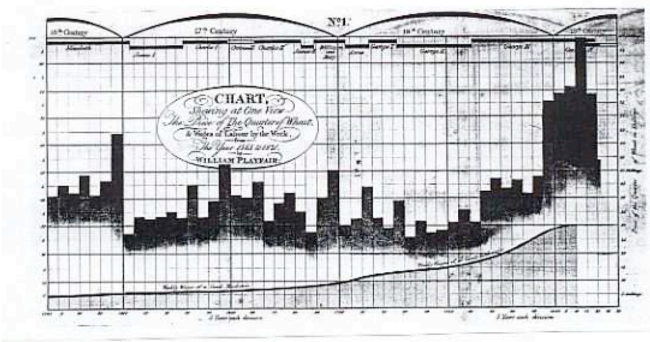

Fig. 1.2: Playfair (1821): Three parallel time series of wheat prices, wages, and the reigns of British kings and queens. From E.R. Tufte (1983).

- E.W Maunder's 1904 butterfly diagram shows a distributional cycle of sunspots moving from the centre of each hemisphere towards the equator, as Galileo had noted in the early 1700's. The first butterfly diagram in Figure 1.3 gives a strong visual measure of variation about the average.

- The modern butterfly diagram increases the data density 10-fold, shows a full century of solar memoirs, 9 cycles of sunspots (see second butterfly diagram in Figure 1.3). The time series shows the area of sun covered by sunspots as a measure of sunspot activity (obtained by summing over all latitudes at any given time). This display is called *parallel-sequencing*.

  - Yule introduced autoregressive processes;

  - he proposed methods for investigating irregular amplitudes, distances between successive peaks and troughs, and

  - suggested incorporating 'shocks' into the future behaviour of the series: stochastic processes.

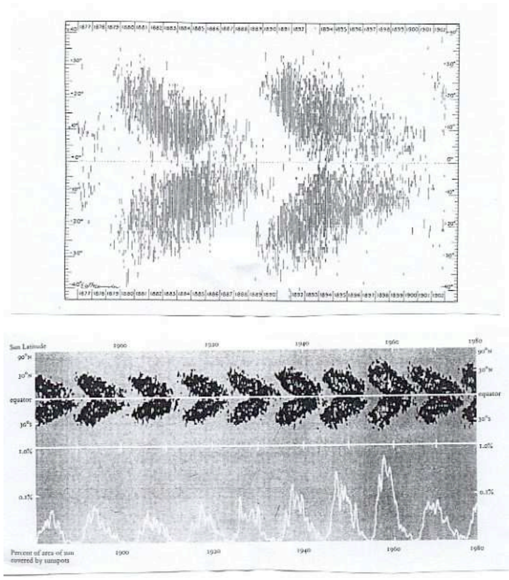

Fig. 1.3: Sunspot data: from E.R. Tufte (1990).

Further examples

We will discuss these examples in the lectures.

1. Newcomb's measurement of the speed of light (deviations from 24,800 nanoseconds): an example of a simple time plot (Figure 1.4).

2. Monthly sales of Australian red wine (Figure 1.5).

3. Biweekly measles notifications and yearly births for 60 cities in England and Wales from 1944 to 1964 (Figure 1.6).

4. The value of the Australian dollar.

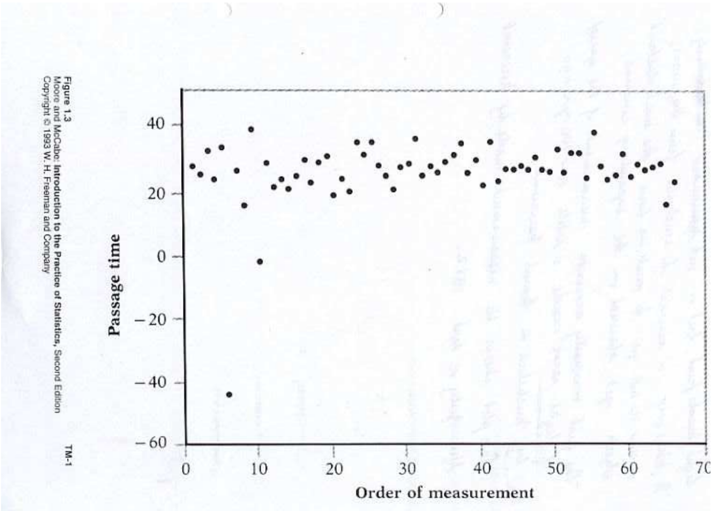

Fig. 1.4: Time plot of Newcomb's measurements of the speed of light.

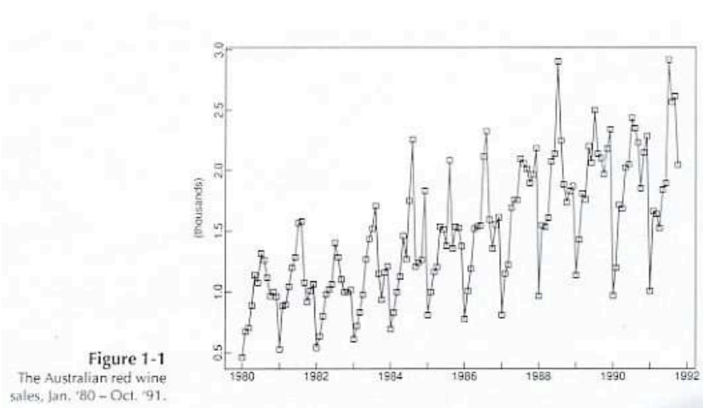

Fig. 1.5: Monthly sales (in kilolitres) of red wine by Australian wine-makers from January 1980 to October 1991. From Brockwell & Davis, 1996.

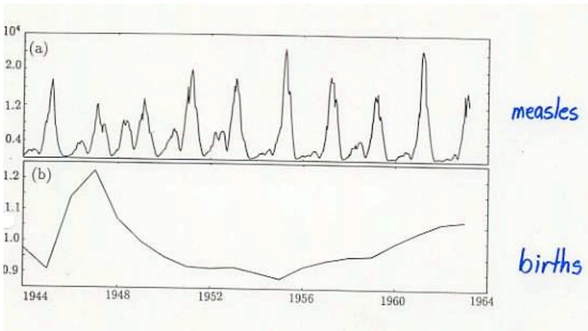

Fig. 1.6: (a) Biweekly measles notifications and (b) yearly births for 60 cities in England and Wales 1944–1964. From Finkenstädt & Grenfell *Applied Statistics*, 2000.

1.2 Characteristics of time series

A univariate time series consists of values of a variable recorded over a long period of time, each observation y_t, $t = 1, \ldots, n$, being recorded at a specific time t.

Time series require special methods for analysis because of the presence of *serial correlation*, a form of 'time dependence' between the observations. For example, in a series of hourly blood pressure readings, a high reading at 1pm is likely to have a certain inertia and to remain high at 2pm. This is an example where neighbouring observations in a time series are *positively correlated*.

Moreover, there are always few, and typically only one independent replication of the data from which to estimate this correlation.

Many time series have one or more of the following characteristics:

- **Order of observations is important**. This implies some sort of dependence or **serial correlation**.
- **Existence of a trend**, or an apparent systematic component.
- **Fluctuations** about the trend; these can be regular or irregular.
- **A deterministic** or other **regular cycle** e.g. a *seasonal component*. By *seasonal*, we mean any periodic behaviour known *a priori*.
- **Cyclic fluctuations** or changes of an irregular nature. By *cyclic*, we mean any periodic behaviour which may or may not be known in advance.
- **Residual or random effects.**

In economic applications, interest often lies in the diffusion about the trend or drift rather than the drift itself, e.g., the volatility of the stock market about a general trend.

Modern methods for the analysis of time series can be divided roughly into two classes:

- The *time domain*: focus on the original series in time; methods are based on direct modelling of the lagged relationship between a series and its past.
- The *frequency domain*: look at harmonic 'summaries' of the series; called *spectral analysis*.

We will be considering both of these broad classes of methods, with emphasis on analysis in the time domain.

1.3 General Approach to Time Series Analysis

Plot series and examine the main features of the graph, checking in particular if there is

(a) a trend

(b) a seasonal component

(c) any apparent sharp changes in behaviour

(d) any outlying observations.

- Remove the trend and seasonal components to get *stationary residuals.*

- Choose a model to fit the residuals.

1.4 Objectives of Time Series Analysis

The objective depends on the particular application, and the main ones are:

- Describing the series
- Forecasting
- Comparing two or more series (*transfer function models*; *time series regression*)
- Assessing interventions (*intervention analysis*)
- Comparing treatments
- Modelling with a view to understanding.

Data may be

- continuous or discrete
- **aggregated**
- equally or unequally spaced in time.

1.5 Data and Notation

- Random variables denoted by upper case letters
- realized values of random variables or observations denoted by lower case letters
- a time series $\{y_t : \ t = 1, 2, \ldots, n\}$ is a set of realized values of random variables $\{Y_t : \ t = 1, 2, \ldots, n\}$
- for continuous or unequally spaced series, write
  - $Y(t)$ rather than Y_t
  - $y(t)$ rather than y_t
- study of time series is a study of the random sequence $\{Y_t\}$ or the random functions $\{Y(t)\}$.

1.6 Relationship to random processes

Terminology:

1. 'Time series' means both the **data**, and the **process** of which it is a realization.

2. In practice, all observed series are **finite**, but for the theory it is convenient to regard them as **infinitely extendable**. This relates the study of time series to *infinite random sequences*, $\{Y_t\}$, and *random functions*, $\{Y(t)\}$.

3. For random sequences: t integer.
For random functions: t varies over $\mathcal{R}$. Use 'random process' for both.

A *random sequence* $\{Y_t\}$ can arise in a number of ways:

- The underlying time scale is genuinely discrete in that the phenomenon of interest does not exist at intermediate times.
- An underlying random function $\{X(t)\}$ may be sampled at equally spaced time points, e.g., taking blood samples at hourly intervals.
- An underlying random function may be accumulated over equal time intervals so that

$$Y_t = \int_{t-1}^{t} X(s)ds.$$

  For example, monthly deaths from asthma.

Time series models:

A (discrete) time series model for observed data $\{y_t\}$ in principle specifies the joint distribution (or possibly only the means and covariances) of a sequence of random variables $\{Y_t\}$. Similarly for the continuous case.

A **time series** is said to be **continuous** when observations are made continuously. For example, a switch may be 'on' or 'off' so the random variable Y is binary, but the series itself is continuous.

1.7 Some simple (zero-mean) models

1. *i.i.d. noise*: the simplest model for a time series is one in which there is no trend or seasonal (i.e. periodic) component and the observations are i.i.d. random variables with zero mean.

   We know that for observations $y_1, \ldots, y_n$,

   $$P(Y_1 \leq y_1, \ldots, Y_n \leq y_n) = F(y_1) \ldots F(y_n),$$

   where F is the cumulative distribution function. Also, since the random variables are independent, for $h \geq 1$,

   $$\begin{aligned} P(Y_{n+h} \leq y | Y_1 = y_1, \ldots, Y_n = y_n) \\ = P(Y_{n+h} \leq y). \end{aligned}$$

   This model is not very interesting for forecasting, but it is an important building block for more complex models.

2. *A binary process*: this is an example of i.i.d. noise. Consider the sequence $\{Y_t, t = 1, \ldots\}$ where $P(Y_t = 1) = p$ and $P(Y_t = -1) = 1-p$, where $p = 1/2$. *How could you produce a realization of this process?*

3. *A random walk*: A random walk with zero mean is obtained by defining $S_0 = 0$ and $S_t = Y_1 + Y_2 + \cdots + Y_t$ for $t = 1, 2, \ldots$, where $\{Y_t\}$ is iid noise.

   If $\{Y_t\}$ is the binary process of Example 2 above, then $\{S_t, t = 0, 1, \ldots\}$ is called a simple symmetric random walk.

We will consider some simple examples with trend and seasonality (an example of a cyclic trend) shortly.

1.8 Trend, serial dependence and stationarity

These concepts are central to understanding the *probability structure* of time series data. Assume we are dealing with random functions; the ideas apply also to random sequences.

- **Trend**: $\mu(t) = \mathsf{E}\,[Y(t)]$

- **Serial dependence**: $Y(t)$ and $Y(s)$ are statistically dependent for at least some pairs of times (t, s), $t \neq s$.

  Typically, $Y(t)$ and $Y(s)$ are correlated, and this leads us to define the

  – *autocovariance function of* $\{Y(t)\}$ as

  $$\gamma(t,s) = \mathsf{E}[\{Y(t) - \mu(t)\}\{Y(s) - \mu(s)\}]$$

  (*auto* means *self, own, of or by oneself*)

- **Stationarity**

  Stationarity implies that the probability structure of the time series doesn't change over time, i.e., $\{Y(t)\}$ looks the same at whatever point we begin to observe it.

  – The strongest form of stationarity is called *strict (or full) stationarity*: this requires that the joint distribution of $Y(t_1), \ldots, Y(t_m)$ be identical to that of $Y(t_1 + \tau), \ldots, Y(t_m + \tau)$ for all m and all τ between $-\infty < \tau < \infty$.

  If $m = 1$, the distribution of $Y(t)$ must be the same for all t so that $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$ are both constants which do not depend on the value of t. Furthermore, if $m = 2$, the joint distribution of $Y(t_1)$ and $Y(t_2)$ depends only on $(t_2 - t_1)$ which is callled the *lag*.

  The acf $\gamma(t_1, t_2)$ also depends only on $(t_2 - t_1)$ and may be written as

  $$\gamma(\tau) = E[Y(t) - \mu)(Y(t + \tau) - \mu)].$$

Strict stationarity is a strong but usually uncheckable assumption. In practice, simpler and weaker forms of stationarity are used.

- *Stationarity in mean*: requires that

$$E[Y(t)] = \mu(t) = \mu,$$

  i.e. the mean does not depend on time t.

- *Marginal stationarity*: requires that the marginal distribution of $Y(t)$ does not depend on t.

- *Second-order or weak stationarity*: requires only that the mean is constant and its acf depends only on the lag, so that

$$\mu(t) = \mu, \qquad \gamma(t, s) = \gamma(|t - s|)$$

  No assumptions are made about higher moments than those of second order. Note that the above definition implies that the variance is constant as well.

From now on, unless stated otherwise, stationary will mean second-order stationary.

Example: the simplest example of a stationary series is *white noise*. This consists of a sequence of mutually independent random variables, each with mean zero and finite variance σ^2.

As an exercise, show that the autocovariance function is

$$\gamma(t, s) = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

Note that white noise implies i.i.d. noise, but not vice versa.

Example: Suppose $\{Y(t_1), \ldots, Y(t_n)\}$ is multivariate normal for all $t_1, \ldots, t_n$. The mvn distribution is completely characterized by its first and second moments, and hence by $\mu(t)$ and $\gamma(|t - s|)$, and so it follows that second-order stationary implies strict stationarity for normal processes.

For processes which are very 'non-normal', μ and $\gamma(|t - s|)$ may not adequately describe the process.

Some comments

The notion of 'independent replications' that forms the basis of most statistical models is replaced in time series analysis by the assumption of stationarity. This assumption of 'homogeneity' over time provides some degree of replication within a single time series, thereby making formal inference possible.

Most of the probability theory of time series is concerned with *stationary time series*, so we often need to turn a non-stationary series into a stationary one in order to use this theory. For example, we may remove the trend and seasonal variation from a set of data and then attempt to model the variation in the residuals by means of a stationary stochastic process.

These definitions are for *univariate* time series. They can be generalized to multivariate time series considered simultaneously.

A statistical model

A useful theoretical framework for handling a wide range of practical problems is a model of the form

$$Y(t) = \mu(t) + U(t)$$

where $Y(t)$ is a measurement (or a transformation of a measurement) made at t, $\mu(t)$ is a non-random trend function, and $U(t)$ is a stationary random function with

$$\mathsf{E}\left[U(t)\right] = 0, \qquad \mathsf{E}\left[U(t)U(s)\right] = \gamma(|t-s|).$$

Example: A simple random walk. We will work through this example in the lecture.#

The difficulty in practice is the separation of the fixed part, $\mu(t)$, and the stochastic part, $U(t)$, and inevitably this involves assumptions. For example, by proposing a parametric model for $\mu(t)$ or $U(t)$, or by assuming the trend is 'smooth', while the random component is 'rough'.

Note however, that judgement of what is rough and what is smooth can depend on the *scale* of observation.

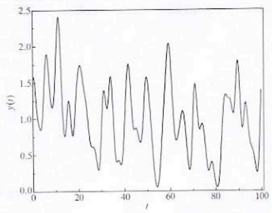

Fig. 1.12. A simulated realization of a stationary random function on the interval $0 \leq t \leq 100$.

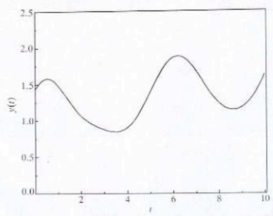

Fig. 1.13. Part of the *same* realization of the stationary random function shown in Fig. 1.12, but restricted to the interval $0 \leq t \leq 10$.

Fig. 1.7: A stationary random function, $Y(t) = \sum_{i=1}^{\infty} f(t - T_i)$, where f is the standard normal density, and the T_i form a homogeneous Poisson point process with unit rate. From Diggle (1990).