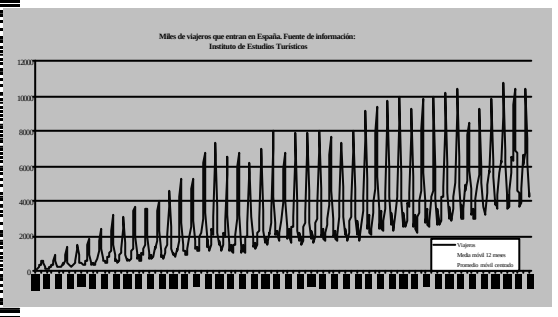
Time Series: Analysis and Forecasting

Josep Allepus

Benevento, May 3rd 2004

Viajeros que entran en España (1958-2000)



Visitor entries

THE SPANISH INBOUND TOURISM SURVEY (FRONTUR)

Visitor entries broken down by type. Serie to March 2004: Provisional data (*)

Visitors Tourists Same-day visitors

		Visitors	Tourists	Same-day
	January	2002 3,925,785	2,263,363	1,662,423
	February	2002 4,424,840	2,656,400	1,768,440
	March	2002 5,784,979	3,795,350	1,989,628
	April	2002 6,039,130	3,814,244	2,224,886
	May	2002 6,789,424	4,647,010	2,142,414
	June	2002 7,131,022	5,110,786	2,020,236
	July	2002 9,869,773	6,862,325	3,007,447
	August	2002 12,199,270	7,824,509	4,374,762
	September	2002 7,629,413	5,287,799	2,341,614
	October	2002 6,528,664	4,473,835	2,054,829
	November	2002 4,720,100	2,800,064	1,920,036
	December	2002 4,981,967	2,791,082	2,190,885
	January	2003 4,279,476	2,436,930	1,842,546
	February	2003 4,383,078	2,698,242	1,684,837
	March	2003 5,488,205	3,344,836	2,143,368
	April	2003 6,712,565	4,283,160	2,429,405
	May	2003 7,378,651	4,864,122	2,514,529
	June	2003 7,510,339	5,122,183	2,388,156
	July	2003 10,117,049	6,932,115	3,184,933
	August	2003 11,847,434	7,294,382	4,553,052
	September	2003 7,652,806	5,140,262	2,512,544
	October	2003 6,791,204	4,478,839	2,312,364
	November	2003 4,907,710	2,800,163	2,107,547
	December	2003 5,358,370	2,917,189	2,441,181
Total	year	2004 15,259,535	8,937,273	6,322,261
	January	2004 4,673,908	2,606,122	2,067,785
	February	2004 5,086,671	2,982,161	2,104,510
	March	2004 5,498,956	3,348,990	2,149,966

Time Series: Analysis and Forecasting

Josep Allepus, URV Spain

http://www.euto.urv.es/estadistica/
/Benev/time_series.pdf

http://www.iet.tourspain.es

Time Series: Analysis and Forecasting

MINISTERIO DE ECONOMÍA
SECRETARÍA DE ESTADO DE COMERCIO Y TURISMO
SECRETARÍA GENERAL DE TURISMO
INSTITUTO DE ESTUDIOS TURÍSTICOS
ESPAÑA

THE SPANISH INBOUND TOURISM SURVEY (FRONTUR)

Tourism entries broken down by type. February 2004 Provisional data (*)

	Total	Month-to-month variation rate	Year-to-year variation rate	Cumulative since January	Cumulative year-to-year variation rate
Visitors	**5,086,671**	**88**	**16.1**	**9,760,579**	**12.7**
Tourists	**2,982,161**	**14.4**	**10.5**	**5,588,284**	**88**
Same-day visitors	**2,104,510**	**18**	**24.9**	**4,172,295**	**18.3**

(*) The figures are definitively reviewed after one year.
Source: IET. - Instituto de Estudios Turísticos.

Cross-sectional and Longitudinal Data

Cross-sectional

- Successive independent samples at the same time.
- The researchers are interested in characteristics of the population at just that point in time.

Longitudinal

- Researchers are interested in how the characteristics of a population may change over time.
- Trend Study - Same group over long period of time.

⊃ A time series is a collection of data over a period of time.

Introduction

- A time series is a set of observations generated sequentially in time
- The observations from a discrete time series,
  - made at some fixed interval h,
    Daily, weekly, monthly, quarterly, ….
  - at times $\tau_1, \tau_2, \ldots, \tau_N$ t = 1,2,…,n
  - may be denoted by $Y(\tau_1), Y(\tau_2), \ldots, Y(\tau_N)$

Introduction (cont.)

Characteristics of time series:

- Time periods are of equal length
- No missing values

- Continuous vs. discrete time series

  Discrete time series may arise in two ways:

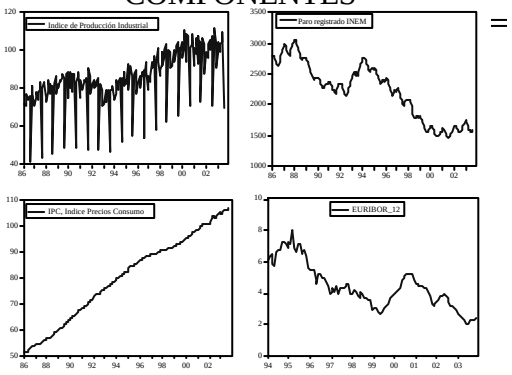
  1. By sampling a continuous time series
  2. By accumulating a variable over a period of time

Time Series in Business and Economics

Very common, applications:

- Economic and business planning
  GDP, Exchange rates …..
- Inventory and production control
- Control and optimization of industrial processes

SERIES TEMPORALES: Ejemplos COMPONENTES



The emphasis in time series is on analysis and forecasting

- First, we examine some of the techniques used in analysing data.
- Finally, we project future events.

  An analysis of h
  be used by mana
  decisions and fo
  planning. We us
  will continue in

DESCRIPCIÓN / INFORMACIÓN

El objetivo es "estudiar" las características de la serie. Puede (debe) combinarse con otras herramientas (autocorrelaciones, análisis gráfico ,tasas de crecimiento,…)

EXPLICACIÓN

El objetivo principal es explicar el "porqué" de la evolución de la serie

The emphasis in time series is on analysis and forecasting

- First, we examine some of the techniques used in analysing data.

PREDICCION

El objetivo principal es obtener predicciones de la serie. Previamente es necesario estudiar las características de la serie.

- Finally, we project future events.
  An analysis of history — a time series — can be used by management to make current decisions and for long-term forecasting and planning. We usually assume past patterns will continue into the future.

Forecasting

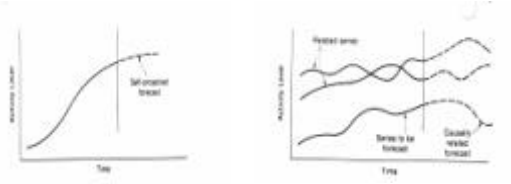
- Lead time of the forecasts
  is the period over which forecasts are needed
- Degree of sophistication
  - Simple ideas
    - ✓ Moving averages
    - ✓ Simple regression techniques
  - Complex statistical concepts: Box-Jenkins methodology

Autocorrelation

- Observations at successive time points are not random, but correlated with each other
- Can use this correlation to develop models for estimating and forecasting time series data

Approaches to forecasting

- Self-projecting approach
- Cause-and-effect approach



Approaches to forecasting (cont.)

- Self-projecting approach
  - Advantages
    - ✓ Quickly and easily applied
    - ✓ A minimum of data is required
    - ✓ Reasonably short-to medium-term forecasts
    - ✓ They provide a basis by which forecasts developed through other models can be measured against
  - Disadvantages
    - ✓ Not useful for forecasting into the far future
    - ✓ Do **not** take into account external factors
- Cause-and-effect approach
  - Advantages
    - ✓ Bring more information
    - ✓ More accurate medium-to long-term forecasts
  - Disadvantages
    - ✓ Forecasts of the explanatory time series are required

Some traditional self-projecting models

- Overall trend models
  - The trend could be linear, exponential, parabolic, etc.
  - A linear Trend has the form $\text{Trend}_t = a + b \cdot t$
  - Short-term changes are difficult to track
- Smoothing models
  - Respond to the most recent behavior of the series
  - Employ the idea of weighted averages
  - They range in the degree of sophistication
  - The simple exponential smoothing method:

$$z_t = \overbrace{A\,z_{t-1} + (1-A)F_{t-1}}^{F_t} + a_t$$

Some traditional self-projecting models (cont.)

- Seasonal models
  - Very common
  - Most seasonal time series also contain long- and short-term trend patterns
- Decomposition models
  - The series is decomposed into its separate patterns
  - Each pattern is modeled separately

Drawbacks of the use of traditional models

- There is no systematic approach for the identification and selection of an appropriate model, and therefore, the identification process is mainly trial-and-error
- There is difficulty in verifying the validity of the model

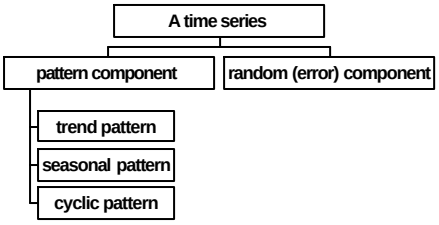
  Most traditional methods were developed from intuitive and practical considerations rather than from a statistical foundation
- Too narrow to deal efficiently with all time series

Forecasts are not always correct

The reality is that a forecast may just be a best guess as to what will happen.
What are the reasons forecasts are not correct? One expert lists eight common errors:

- failure to carefully examine the assumptions,
- limited expertise,
- lack of imagination,
- neglect of constraints,
- excessive optimism,
- reliance on mechanical extrapolation,
- premature closure, and
- overspecification.

Components of a time series



Four Primary Components of a Time Series:

- Secular Trend: The trend is the long-run direction of the time series.
- Seasonal variation is the pattern in a time series within a year. These patterns tend to repeat themselves from year to year for most businesses.
- Cyclical Movements the fluctuation above and below the long-term trend line.
- The Irregular variation is divided into two components:

1. The episodic variations are unpredictable, but they can usually be identified. A flood is an example.
2. The residual variations are random in nature.

Mathematical Representations

Una serie temporal Y(t), puede admitir una descomposición del tipo:

Additive: $Y = T + S + C + I$

Multiplicative: $Y = T * S * C * I$

Esquema mixto

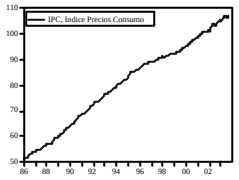
$Y_t = T_t + C_t + E_t + I_t$

Observations:

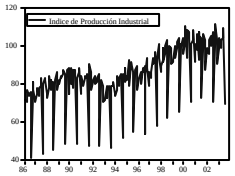
- Traditional time series analysis is "atheoretic". No economic theory guides us in writing down this decomposition.
- Typically, one of these components will dominate and this will affect the behavior of the series.

¿Características dominantes de la serie?

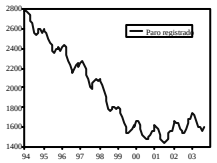
Example: Secular Trend

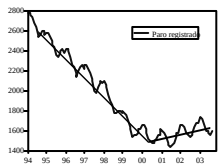


Example: Seasonal Component



Example: Cyclical Component

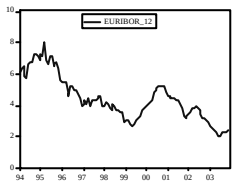




Cyclical Variation

- The rise and fall of a time series over periods longer than one year.
  A typical business cycle consists of a period of prosperity followed by periods of recession, depression, and then recovery.
  There are sizable fluctuations unfolding over more than one year in time above and below the secular trend.

Example: Random/Irregular



Irregular or Random Components

- Special events that pull macro variables off their usual paths.
- Can be expected or unexpected.

Many analysts prefer to subdivide the **irregular variation** into *episodic* and *residual* variations.

- ✓ Episodic fluctuations are unpredictable, but they can be identified. The initial impact on the economy of a major strike or a war can be identified, but a strike or war cannot be predicted.
- ✓ After the episodic fluctuations have been removed, the remaining variation is called the residual variation. The residual fluctuations, often called chance fluctuations, are unpredictable, and they cannot be identified.

Of course, neither episodic nor residual variation can be projected into the future.

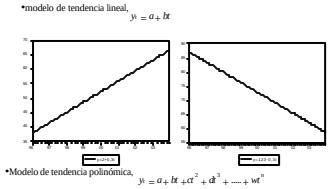
Secular Trends

- Often called "Time Trends"
- Visual representation is called "Time Path" or "Time Shape"
- A continuous set of integers is used to represent time in these models. t = 1,2,…,n
- Linear Time Trend Model

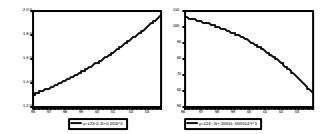
$$Y_t = a + b \cdot t$$

Trend Shapes

•modelo de tendencia lineal, $y_t = a + bt$

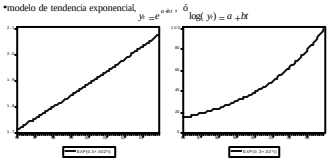


•Modelo de tendencia polinómica, $y_t = a + bt + ct^2 + dt^3 + \cdots + wt^n$

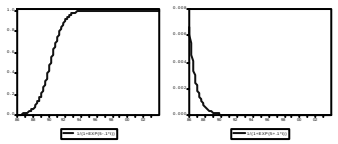


Trend Shapes (2)

•modelo de tendencia exponencial, $y_t = e^{a+bt}$, ó $\log(y_t) = a + bt$



•Modelo de tendencia logística, $y_t = \frac{L}{1+e^{a+b}}$



Determine a linear trend equation

Linear Trend: The long-term trend of many business series, such as sales, exports, and production, often approximates a straight line. If so, the equation to describe this growth is:

LINEAR TREND EQUATION $Y = a + bt$

Y is the projected value of the Y variable for a selected value of t.

a is the Y-intercept. It is the estimated value of Y when $t = 0$. Another way to put it is: a is the estimated value of Y where the line crosses the Y-axis when t is zero.

b is the slope of the line, or the average change in Y for each change of one unit (either increase or decrease) in t.

t is any value of time that is selected.

To simplify the calculations, the years are replaced by *coded* values. That is, we let 1991 be 1, 1992 be 2, and so forth.

Determine a linear trend equation

We drew a line through points on a scatter diagram to approximate the regression line.

The **least squares method** of computing the equation for a line through the data of interest gave the ''best-fitting'' line.

Normal equations: Two equations may be solved simultaneously to arrive at the least squares trend equation. They are:

EQUATIONS FOR THE TREND LINE

$$\Sigma Y = n \cdot a + b\Sigma t$$

$$\Sigma t \cdot Y = a\Sigma t + b\Sigma t^2$$

Nonlinear Trends

A linear trend equation is used to represent the time series when it is believed that the data are increasing (or decreasing) by *equal amounts*, on the average, from one period to another.
Data that increase (or decrease) by *increasing amounts* over a period of time appear *curvilinear* when plotted.
Business series, such as automobile sales, shipments of soft-drink bottles, and residential construction, have periods of above-average and below-average activity each year.

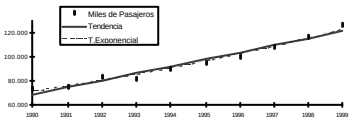
Determine a trend

Viajero91.xls

Estimación de la tendencia por el método de mínimos cuadrados para miles de Viajeros por Transporte aéreo en España

Año	Pasajeros	X	Tendencia	Log(Y)	X	Log(Y²)	Tend.Log
1990	73.143	1	68.709	11,20	1	11,17	71.051
1991	75.233	2	74.549	11,23	2	11,23	75.622
1992	82.672	3	80.389	11,32	3	11,29	80.273
1993	81.401	4	86.229	11,31	4	11,35	85.324
1994	89.498	5	92.069	11,40	5	11,42	90.692
1995	95.432	6	97.909	11,47	6	11,48	96.398
1996	100.722	7	103.749	11,52	7	11,54	102.463
1997	108.623	8	109.588	11,60	8	11,60	108.910
1998	116.806	9	115.428	11,67	9	11,66	115.762
1999	126.356	10	121.268	11,75	10	11,72	123.046

Tendencias lineal y exponencial por MCO para la serie de Viajeros por Transporte aéreo en España



How predictable is the secular trend in a series?

Removing Time Trends: Detrending

Often, the trend component of a time series dominates, but the interesting part of the series is another component.

Detrending

- Step 1: Estimate the *Secular Trend* using regression model
- Step 2: Subtract the estimated secular trend from the original series.

Note: This is also the "Residual Approach" to analyzing cyclical data

Seasonal Component

- Found in High Frequency data (Quarterly, monthly)
- Caused by natural or budget calendars
  - Retail Sales higher during holidays
  - Travel more frequent in summer
  - Weather
- Want to quantify or remove in forecasting
- How predictable is this component?

Seasonal Variation

Patterns of change in a time series within a year. These patterns tend to repeat themselves each year.
Almost all businesses tend to have recurring seasonal patterns.

Compute a moving average

The Moving-Average Method : *Moving-average method smooths out fluctuations*

The **moving-average method** is not only useful in smoothing out a time series; it is the basic method used in measuring the seasonal fluctuation, described.

In contrast to the least squares method, which expresses the trend in terms of a mathematical equation ($Y = a + bt$), the moving-average method merely smooths out the fluctuations in the data. This is accomplished by ''moving'' the arithmetic mean values through the time series.

To apply the moving-average method to a time series, the data should follow a fairly linear trend and have a definite rhythmic pattern of fluctuations. If the duration of the cycles is constant, and if the amplitudes of the cycles are equal, the cyclical and irregular fluctuations can be removed entirely using the moving-average method. The result is nearly a straight line.

The first step in computing the twelve-months moving average is to determine the twelve-months moving totals and determine the arithmetic mean sales per year.

Método de la Media Móvil:

Se basa en el suavizado de la serie mediante medias móviles sucesivas de orden "p".
PASOS:
1. Representar gráficamente la serie y observar cuál es el período de oscilaciones más importantes.
2. Elegir un valor de "p" que represente el período de oscilaciones más importantes que caracteriza la serie (m.c.m. de ciclo y estacionalidad).
Si "p" es par las medias móviles serían:
por lo que sería necesario centrarlas haciendo la media de medias móviles sucesivas:
La tendencia de la serie la componen las medias móviles centradas obtenidas en el paso 2.

MEDIAS MÓVILES

Una media móvil no es más que el valor medio de un conjunto de valores adyacentes de una serie temporal, existiendo dos tipos genéricos: medias móviles simétricas o centradas y medias móviles asimétricas.

Una media móvil simétrica de la variable y_t de orden 2p+1, denotada por MM(2p+1), es aquella que contiene p términos anteriores y p términos posteriores de la variable y_t, y viene dada por la expresión:

$$MM(2p+1)_t = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_t + y_{t+1} + y_{t+2} + \ldots + y_{t+p}}{2p+1}$$

Una media móvil asimétrica es aquella que no cuenta con un conjunto simétrico de valores de yt, en concreto, una media móvil asimétrica de orden p responde generalmente a la expresión:

$$MMA(p) = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_t}{p}$$

Determining a Seasonal Index

Objective: To determine a set of "typical" seasonal indexes
A typical set of monthly indexes consists of 12 indexes that are representative of the data for a 12-month period. Logically, there are four typical seasonal indexes for data reported quarterly. Each index is a percent, with the average for the year equal to 100.0; that is, each monthly index indicates the level of sales, production, or another variable in relation to the annual average of 100.0.
A typical index of 96.0 for January indicates that sales (or whatever the variable is) are usually 4 percent below the average for the year.
An index of 107.2 for October means that the variable is typically 7.2 percent above the annual average.
Several methods have been developed to measure the typical seasonal fluctuation in a time series. The method most commonly used to compute the typical seasonal pattern is called the **ratio-to-moving-average method.** It eliminates the trend, cyclical, and irregular components from the original data (Y).
The numbers that result are called the *typical seasonal index.*

Índices de Variación Estacional

La estacionalidad de cada período vendrá representada por los (IGVEAk) correspondientes a cada uno de los períodos.
Cálculo de los Índices Específicos de Variación Estacional (IEVEik) según el esquema de acuerdo con el que se combinan las componentes de la serie sea:

1. Aditivo
2. Multiplicativo

INTERPRETACIÓN:

En el esquema aditivo: Cuando un Índice General de Variación Estacional Ajustado sea positivo, entonces la variable supera a la media de tendencia-ciclo en dicho período, debido al efecto estacional; dándose el efecto contrario si es negativo.

En el esquema multiplicativo: Cuando un Índice General de Variación Estacional Ajustado es mayor que 1 (que 100 en %), entonces la variable supera a la media de tendencia-ciclo en dicho período, por el efecto estacional; y viceversa si es menor que 100%.