Module 1

Introduction to Time Series

Class notes for Statistics 451: Applied Time Series
Iowa State University

Copyright 2015 W. Q. Meeker.

November 11, 2015
17h 19min
Module 1

Segment 1

Time Series Examples
Time Series Data

A sequence of observations taken over time (usually equally spaced) \( y_1, y_2, \ldots, y_t, \ldots y_n \) where \( n \) is the number of observations in the “realization”

- **Univariate** – single series, e.g., daily closing price of IBM common stock.
- **Multivariate** – two or more series (vector time series) (e.g. daily closing price of common stock of Apple, IBM, Microsoft gives 3 values per day).
- **Time interval**: yearly, monthly, quarterly, weekly, daily, hourly, every minute, … every \( k \) nanoseconds …. +
- **Population of units** → **Sample**
- **Process operating in time** → **Realization**
Continuous Time and Discrete Time

• Time is continuous, but data are usually reported at discrete points in time.

• Thus “sampling” a continuous time series leads to a discrete time series.

• Sampling is usually equally spaced in time (sampling sometimes known as “reporting”)

• Time series data are usually not independent, especially if sampling interval is small. Observations close together are often more alike than those far apart (e.g. daily temperatures).
Manufacturer’s Shipments 1967-1975

![Graph showing manufacturer's shipments from 1967 to 1975. The graph plots ship.ts against time (1968, 1970, 1972, 1974). The shipments show a general upward trend with fluctuations.](plot(ship.ts))
Smoothed Manufacturer’s Shipments 1967-1975

plot(ship.ts)
lines(smooth(ship.ts), col=2, lwd=3)
Seasonal Data Example
US Housing Starts 1966-1974

US Housing Starts 1966−1974

Thousands of Homes

Year

Types of Time Series Responses

- Continuous (e.g. temperature, concentration).
- Discrete (e.g., number of people, number manufactured) (often from aggregation). We often approximate discrete responses with a continuous model.
- Binary (e.g., success or failure).
Time Series Applications

- Economics (e.g., GNP, NNP, Unemployment Rate, Interest Rates, Money supply)
- Business (e.g., Inventory, Cash, Sales, Prices, Quality Indices, Stock Price)
- Sociology (e.g., Crime Rates, Divorce Rates)
- Meteorology (e.g., Rainfall, Temperature, Wind Speed)
- Astronomy (e.g., Solar Activity, Sun Spots, Star Brightness)
- Ecology (e.g., Air Pollution, Water Pollution, Wildlife Population)
- Engineering
International Airline Passengers 1949-1960

International Airline Passengers

Year

Thousands of Passengers


100 200 300 400 500 600
Log International Airline Passengers 1949-1960

Log International Airline Passengers

log(Thousands of Passengers)

Year


5.0 5.5 6.0 6.5

log(Thousands of Passengers)

Year


5.0 5.5 6.0 6.5
Reasons for Analyzing Time Series

• Description of features (level, trend, cycles, seasonality, variability)
  ▶ Graphical (time series plots)
  ▶ Numerical (mean, standard deviation, autocorrelation)

• Explanation
  Explanatory Variables

• Prediction (Forecasting)
  Next Year’s Sales

• Control
  Quality of Manufacturing Process Economy
Module 1

Segment 2

Time Series Filters
Filters

- Filters are like functions, but for time series. Let $x = x_1, x_2, \ldots$ and $y = y_1, y_2, \ldots$. Then $y = f(x)$:

  \[ y_t = f(x_t) \]

- Linear filter:

  \[ y_t = \sum_{r=-q}^{+s} a_r x_{t+r} = a_{-q} x_{t-q} + a_{-q+1} x_{t-q+1} + \ldots + a_0 x_t + \ldots + a_s x_{t+s}. \]

  For example with $q = 2$, $s = 2$, and $a_r = 1/5$,

  \[ y_t = \sum_{r=-2}^{2} a_r x_{t+r} = (x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})/5 \]

  which is a “moving average” filter.
Filters in Series

Two (or more) filters in series form an overall filter:

\[
x_t \xrightarrow{f_1} y_t \xrightarrow{f_2} z_t
\]

<table>
<thead>
<tr>
<th>(x_1)</th>
<th>(x_1 + x_2)/2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_2)</td>
<td>((x_1 + 2x_2 + x_3)/4)</td>
</tr>
<tr>
<td>(x_3)</td>
<td>((x_2 + 2x_3 + x_4)/4)</td>
</tr>
<tr>
<td>(x_4)</td>
<td>((x_3 + x_4)/2)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

\[
z_t = \sum_{r=-1}^{1} a_r x_{t+r}
\]

where \(a_{-1} = 1/4\), \(a_0 = 1/2\), and \(a_1 = 1/4\). This particular weighted moving average is also known as a “Hanning filter.”
Smoothed International Airline Passengers 1949-1960
12-point moving average

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousands of Passengers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
<td>100</td>
</tr>
<tr>
<td>1951</td>
<td>200</td>
</tr>
<tr>
<td>1953</td>
<td>300</td>
</tr>
<tr>
<td>1955</td>
<td>400</td>
</tr>
<tr>
<td>1957</td>
<td>500</td>
</tr>
<tr>
<td>1959</td>
<td></td>
</tr>
<tr>
<td>1961</td>
<td></td>
</tr>
</tbody>
</table>
Smoothed International Airline Passengers 1949-1960
(4253H,twice)
Difference Noise
International Airline Passengers 1949-1960

\[
t\textsf{tsplot(passengers.ts-smooth.spline(passengers.ts))}
\]

Passenger-smooth(Passenger)
Ratio Noise

International Airline Passengers 1949-1960

tsp\text{plot}(\text{passengers.ts}/\text{smooth}\text{.spline(\text{passengers.ts}))}

Passenger/\text{smooth(\text{Passenger})}
Module 1

Segment 3

Classical Time Series Decomposition
The “Classical” Decomposition Model

Multiplicative Model:

\[ Z_t = T_t \times C_t \times S_t \times I_t \]

- Decompose \( Z_t \) into \( T_t, C_t, S_t \) and \( I_t, t = 1, \ldots, n \) (trend, cycle, seasonal, and irregular).

- There are many possible algorithms for doing the decomposition.

- Operationally it may not be desirable to separate \( T_t \) and \( C_t \).

- An **additive model** could also be used (especially after a transformation)

\[ Z_t = T_t + C_t + S_t + I_t \]
Seasonal Decomposition of Time Series by Loess
International Airline Passengers 1949-1960
Periodic Seasonality

plot(stl(Passengers.ts^(0.25), s.window="periodic"))
Seasonal Decomposition of Time Series by Loess
International Airline Passengers 1949-1960
Evolving Seasonality

plot(stl(Passengers.ts^(0.25), s.window=7))
Applications of the “Classical” Decomposition Model

Model:

\[ Z_t = T_t \times C_t \times S_t \times I_t \]

- Forecast components into future and combine to forecast \( Z_t \). Suppose there are \( n = 100 \) observations in a realization.

  \( \downarrow \) Forecast \( T_{101}, C_{101}, S_{101}, I_{101} \).

  \( \downarrow \) \( \hat{Z}_{101} = \hat{T}_{101} \times \hat{C}_{101} \times \hat{S}_{101} \times \hat{I}_{101} \)

- Smoothed series:
  \[ \text{Smooth}_t = T_t \times S_t \times C_t \quad (\text{no } I_t) \]

- Seasonally adjusted series (also known as deseasonalized):
  \[ D_t = T_t \times C_t \times I_t \quad (\text{no } S_t) \]
For more information on decomposition methods, see

- Some elementary business statistics textbooks

- Census Bureau program X-12 ARIMA (www.census.gov)

- Census Bureau program X-13 ARIMA SEATS (Signal Extraction ARIMA Time Series Models)

- SAS ETS PROC X-12

- S-Plus function sabl() (Seasonal Analysis Bell Labs)

- R function st1() (Seasonal Decomposition by Loess)
Module 1

Segment 4

Shaded Time Series Plots
US Housing Starts 1966-1974

shaded.tsplot(hstart.ts, top=F)
US Housing Starts 1966-1974

\texttt{shaded.tsplot(hstart.ts,top=T)}
Balance of Trade Between England and NA 1700-1780
William Playfair, Commercial and Political Atlas, 1786
Balance of Trade in England 1700-1780
William Playfair, Commercial and Political Atlas, 1786
Playfair Recreated
http://dougmccune.com/flex/recreatingPlayfair/
Module 1

Segment 5

Introduction to Autocorrelation
Percent $\text{CO}_2$ Outlet Gas
(sampling interval 9 seconds)
plot(gasry.tsd)
Visualization of the Autocorrelation Function for the Percent CO$_2$ Outlet Gas Data

show.acf(gasrx.tsd)
Autocorrelation Function
for the Percent \( \text{CO}_2 \) Outlet Gas Data
\[ \text{acf(gasrx.tsd)} \]

Series : gasry.d$ts
RTSERIES `iden` Command Output
for the Percent CO\textsubscript{2} Outlet Gas Data
`iden(gasrx.tsd)`

CO\textsubscript{2} (output)
w= Percent CO\textsubscript{2}

Range-Mean Plot

ACF

PACF