Time Series

Robert Winslow

GalvanizeU

July 6th, 2017

Review

Lab questions?

Warm up discussion

Summarize the last four lessons.

Warm up discussion

Why do we make assumptions in our models?

- Prior distributions
- Exchangeability (ordering of RVs)

Time series

Univariate time series: a sequence of measurements of the same variable collected over time.

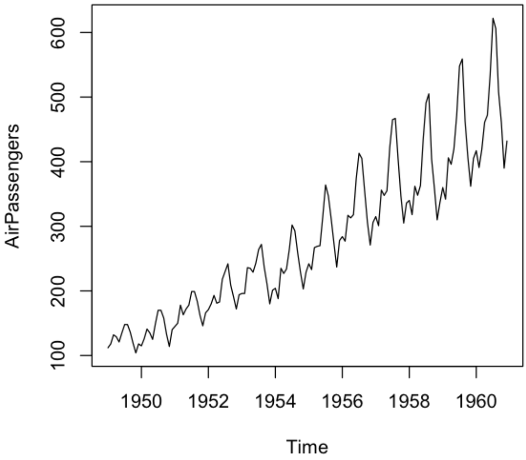
Most often, the measurements are made at regular time intervals.

Time series data

Time series data is a sequence of observations of some quantity of interest, which are collected over time, such as:

- GDP
- The price of gold or a stock
- Demand for a good
- Unemployment
- Web traffic (clicks, logins, posts, etc.)

Time series



AirPassengers
100
200
300
400
500
600
1950
1952
1954
1956
1958
1960
Time

Basic Objectives of the Analysis

The basic objective usually is to determine a model that describes the pattern of the time series. Uses for such a model are:

- To describe the important features of the time series pattern.
- To explain how the past affects the future or how two time series can “interact”.
- To forecast future values of the series.
- To possibly serve as a control standard for a variable that measures the quality of product in some manufacturing situations.

Time series

Basic Methodology

- A "pattern" is first attained from the data at hand.
- The "pattern" is then extrapolated into the future to prepare a forecast.

Assumption

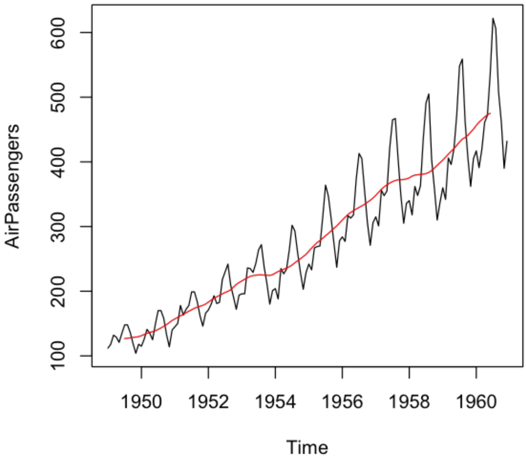
The "pattern" we've observed will continue...

Time series

Some applications:

- Predicting/forecasting future gas prices
- Global warming? Predicting future global temperatures
- Population growth

Time series: trend example



Time series: why special methods?

One difference from standard linear regression is that the data are not necessarily independent and not necessarily identically distributed.

One defining characteristic of time series is that this is a list of observations where the ordering matters.

Ordering is very important because there is dependency and changing the order could change the meaning of the data.

Time series: Types of Models

There are two basic types of “time domain” models.

- Models that relate the present value of a series to past values and past prediction errors - these are called ARIMA models (for Autoregressive Integrated Moving Average).
- Ordinary regression models that use time indices as x-variables. These can be helpful for an initial description of the data and form the basis of several simple forecasting methods.

Time series: Modeling considerations

- Is there a trend, meaning that, on average, the measurements tend to increase (or decrease) over time?
- Is there seasonality, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?
- Are their outliers? In regression, outliers are far away from your line. With time series data, your outliers are far away from your other data.
- Is there a long-run cycle or period unrelated to seasonality factors?
- Is there constant variance over time, or is the variance non-constant?
- Are there any abrupt changes to either the level of the series or the variance?

Time series definition

We assume a time series, $\{y_t\}$, has the following properties:

- y_t is an observation of the level of y at time t
- $\{y_t\}$ is time series, i.e., the collection of observations:
  - May extend back to $t = 0$ or $t = -\infty$, depending on the problem.
  - E.g., $t \in \{0, ..., T\}$
- Starting after time t, we typically want to forecast y out to a "horizon" h, namely forecasting values $y_{t+1}, y_{t+2}, ...y_{t+h}$.

To emphasize our knowledge of y up to time t, this can be written as $y_{t+1|t}, y_{t+2|t}, ...y_{t+h|t}$

Assumptions

- Discrete time:
- Sampling at regular intervals, ...
- even if process is continuous
- Evenly spaced observations
- No missing observations

Caveat: only one observation?

Time series are hard to model because we only observe one realization of the path of the process:

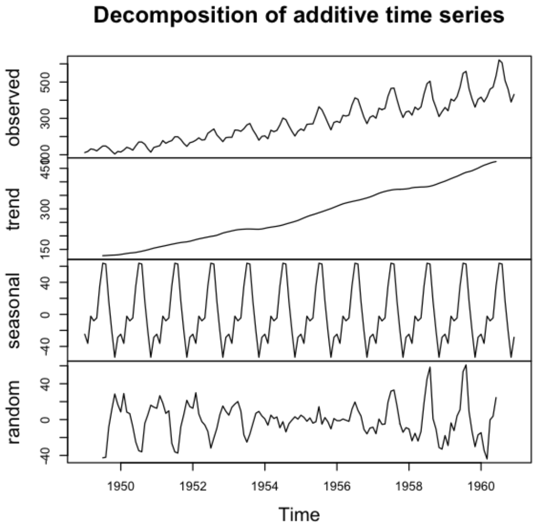
- Often have limited data
- Must impose structure – such as assumptions about correlation – in order to model
- Must project beyond support of the data

Components of a Time Series

- Trend: the long-run upward or downward movement of the series
- Cycle: the upward or downward movement of the series around the trend (think of a wave).
- Seasonal Variations: patterns in the data that follow regular time-based patterns (think of seasonal temperatures).
- Irregular Variations: the remaining erratic movements in the series that cannot be accounted for

Our **goal** is to estimate the trend, cycle, and seasonal components of a time series so that all that is left is irregular fluctuations (often referred to as white noise).

Time series



Decomposition of additive time series
observed
trend
seasonal
random
500
300
100
450
300
150
40
0
-40
40
0
-40
1950
1952
1954
1956
1958
1960
Time

Time Series Decomposition

- An additive model

$$y_t = S_t + T_t + E_t$$

where y_t is the data at period t, S_t is the seasonal component at period t, T_t is the trend-cycle component at period t and E_t is the remainder (or irregular or error) component at period t

- A multiplicative model

$$y_t = S_t \times T_t \times E_t$$

Time Series Decomposition

The additive model is most appropriate if the magnitude of the seasonal fluctuations or the variation around the trend-cycle does not vary with the dependent variable of the time series.

When the variation in the seasonal pattern, or the variation around the trend-cycle, appears to be proportional to the level of the time series, then a multiplicative model is more appropriate.

With economic time series, multiplicative models are common.

Time Series Decomposition

An alternative to using a multiplicative model is to first transform the data until the variation in the series appears to be stable over time, and then use an additive model.

When a log transformation has been used, this is equivalent to using a multiplicative decomposition because

$y_t = S_t \times T_t \times E_t$ is equivalent to
$log(y_t) = log(S_t) + log(T_t) + log(E_t)$.

Sometimes, the trend-cycle component is simply called the "trend" component, even though it may contain cyclic behaviour as well.

Time series

One of the simplest ARIMA type models is a model in which we use a linear model to predict the value at the present time using the value at the previous time.

This is called an AR(1) model, standing for autoregressive model of order 1.

The order of the model indicates how many previous times we use to predict the present time.

Time series

AR(1):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

AR(2):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$$

AR(n):

$$y_t = \beta_0 + \beta_1 y_{t-1} + ... + \beta_n y_{t-n} + \epsilon_t$$