Day 3 - Decomposition of time series

Introduction

The purpose of today's lecture is to understand how to decompose a time series into its constituent components in R. To guide our exploration, we will return to the Vermont temperatures and Pan Am data sets.

```
# packages
library(tidyverse)

# load data and rename
## Pan Am
data(AirPassengers)
ap <- AirPassengers

## vt temps
vt_temps <- readr::read_csv("vt_temps.csv")
```

Review: creating `ts` objects

Create a `ts` object, called `vt_ts`, for the monthly temperatures in Vermont that spans from 1970/06/01 to 2013/04/01. Plot the time series.

```r
# alternative using window()
vt_ts_long <- ts(
  vt_temps$AverageTemperature,
  start = c(1850, 1),
  end = c(2013, 9),
  freq = 12
)

vt_ts <- window(
  vt_ts_long,
  start = c(1970, 6),
  end = c(2013, 4)
)
```

Introducing definitions and notation

> **Random variables**
>
> A ______________________________, usually written X, is a variable whose possible values are the numerical outcomes of ________________ phenomenon. There are two types of random variables, _____________________ and __________________.
>
> Formally, a random variable is a mapping from a sample space S to the real numbers.

Discrete random variables:

Continuous random variables:

> **Time series notation**
>
> A ______________________________ of length n is a sequence of ________________________, which we denote $\{X_t : t = 1, \dots, n\} = \{X_1, X_2, \dots, X_n\}$. When referring to an observed time series, we use lowercase letters, $\{x_t : t = 1, \dots, n\} = \{x_1, x_2, \dots, x_n\}$. If the length of the series n does not need to be specified, we will often use the abbreviated notation $\{x_t\}$.

- $\bar{x} = \frac{\sum x_i}{n}$

- $\hat{x}$

- $\hat{x}_{t+k|t}$

Our first time series model

Decomposition models

An ______________________________ is a simple model for a time series that estimates the ____________, ______________________, and ____________.

$$x_t = m_t + s_t + z_t$$

A ___________________________________ allows for the seasonal effect to increase as the trend increase.

$$x_t = m_t \cdot s_t \cdot z_t$$

If the time series is strictly positive, it may be easier to fit an additive model on the log scale than a multiplicative model on the original scale.

$$\log(x_t) = m_t + s_t + z_t$$

Estimating m_t, s_t, and z_t

How can we obtain an estimate of the trend effect?

Centered moving average

For time series with a period of 12 (i.e. monthly data), the ________________________________ at time t is given by

$$\hat{m}_t = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + \cdots + x_{t-1} + x_t + x_{t+1} + \cdots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}$$

where $t = 7, \ldots, n-6$

How can we obtain an estimate of the seasonal effect at each time t? How can we obtain an estimate of the overall seasonal effect associated with each month?

Seasonal effects

For an additive time series with a monthly frequency, the seasonal effect at time t is estimated by

$$\hat{s}_t = x_t - \hat{m}_t$$

We can obtain a single estimate of the monthly effect by averaging the effect of each month.

$$\bar{s}_t = \frac{\sum s_t}{T-1}$$

where T denotes the number of years. Often times, the estimated seasonal effect is **centered** after calculation - more on this on Wednesday. If a time series is multiplicative, the seasonal effect is instead estimated by

$$\hat{s}_t = \frac{x_t}{\hat{m}_t}$$

How can we obtain an estimate of z_t?

Residual error series

The ____, also called ____________________, is the raw time series adjusted for the trend and seasonal effects. On average, this series should have a mean of ___________________. For an additive decomposition model, the residual error series is

$$\hat{z}_t = x_t - \hat{m}_t - \bar{s}_t$$

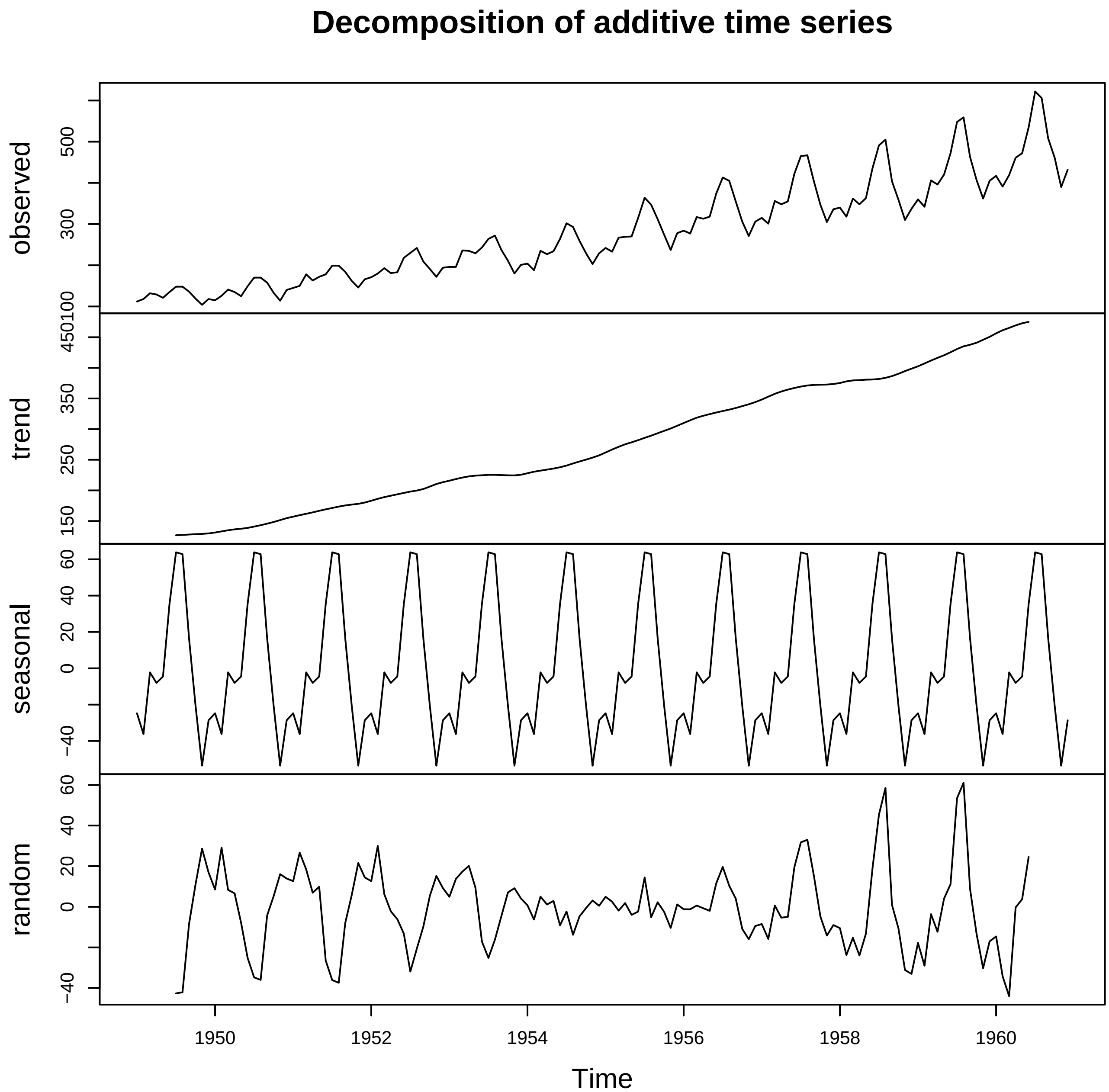
For a multiplicative decomposition model, the residual error series is

$$\hat{z}_t = \frac{x_t}{\hat{m}_t \cdot \bar{s}_t}$$

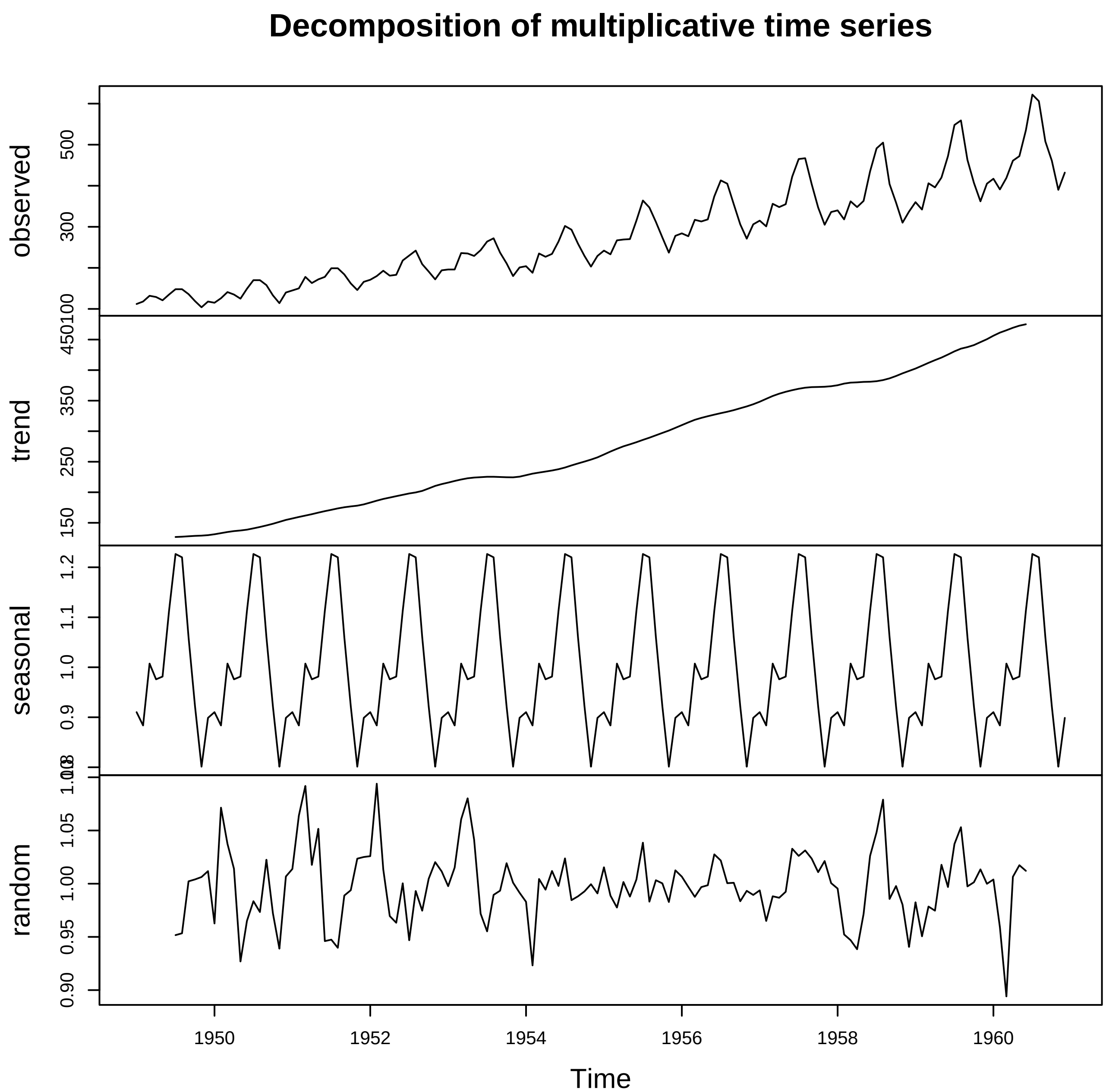
Decomposition in R

> **Note**
>
> The `decompose` function may be used in R to obtain a decomposition of a time series object.

```r
plot(decompose(ap))
```



```
plot(decompose(ap, type = "multiplicative"))
```



```
plot(decompose(log(ap)))
```