Introduction to time series

Lecture notes

2025-01-08

Question: What is time series?

- Data observed at regular time intervals (yearly, monthly, weekly, daily, etc.).
- The time index or interval (yearly, monthly, weekly, daily, etc.) are often called frequnecy of the time series
- Widely used in economics, finance, healthcare, climate science, transport, and engineering.

Examples of time series data;

(a) In Economics: GDP, monthly unemployment rates, inflation rates,stock market prices,sales, prices, etc.

- For example, the data below shows the first and last few rows of the GDP growth rate of the U.S., from the first quarter of 1948 to the first quarter of 2016.

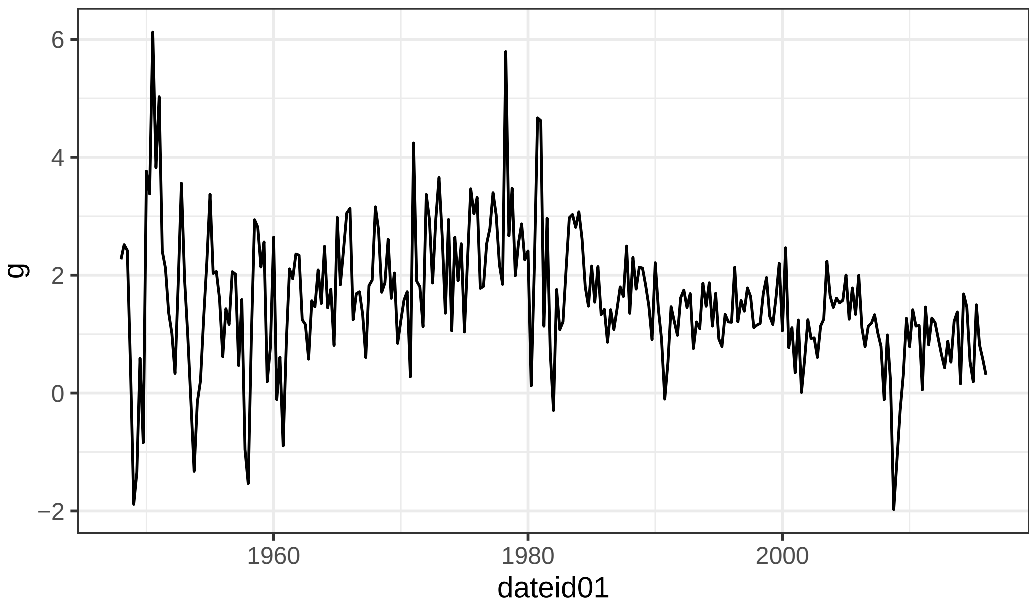
```
    dateid01      g
1 1948-01-01  2.267
2 1948-04-01  2.517
3 1948-07-01  2.418
4 1948-10-01  0.429
5 1949-01-01 -1.888
6 1949-04-01 -1.344
```

```
      dateid01     g
268 2014-10-01 0.535
269 2015-01-01 0.190
270 2015-04-01 1.498
271 2015-07-01 0.818
```

```
272 2015-10-01 0.579
273 2016-01-01 0.310
```

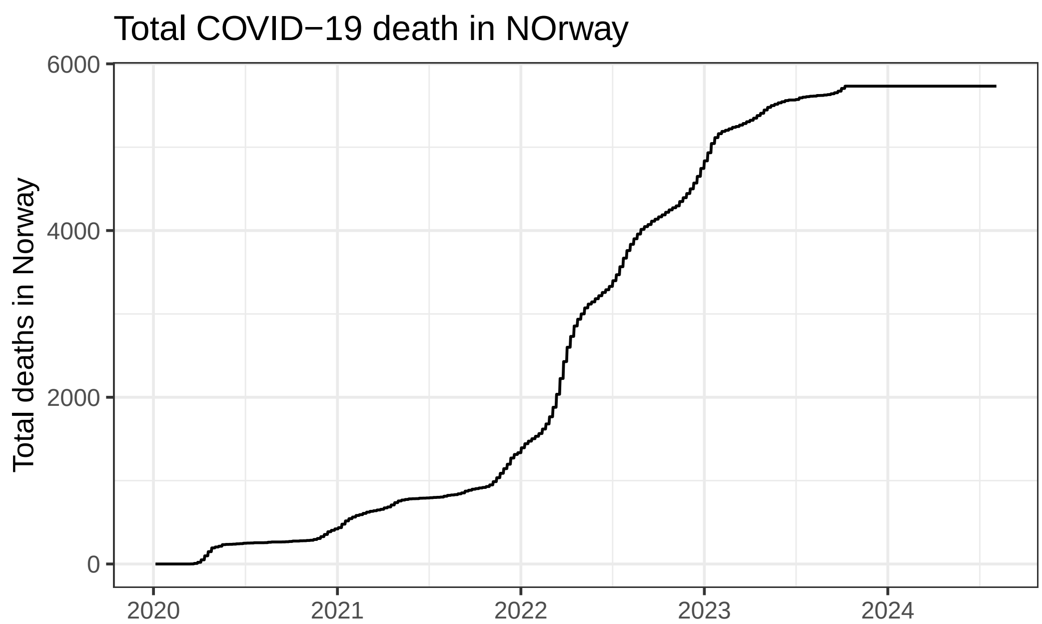
- 273 quarters (69 years) measured from the first quarter of 1948 to the first quarter of 2016.
- When working with time series data, it is essential to understand its structure and characteristics.
- Ploting the data helps to identify patterns, trends, seasonality, etc.

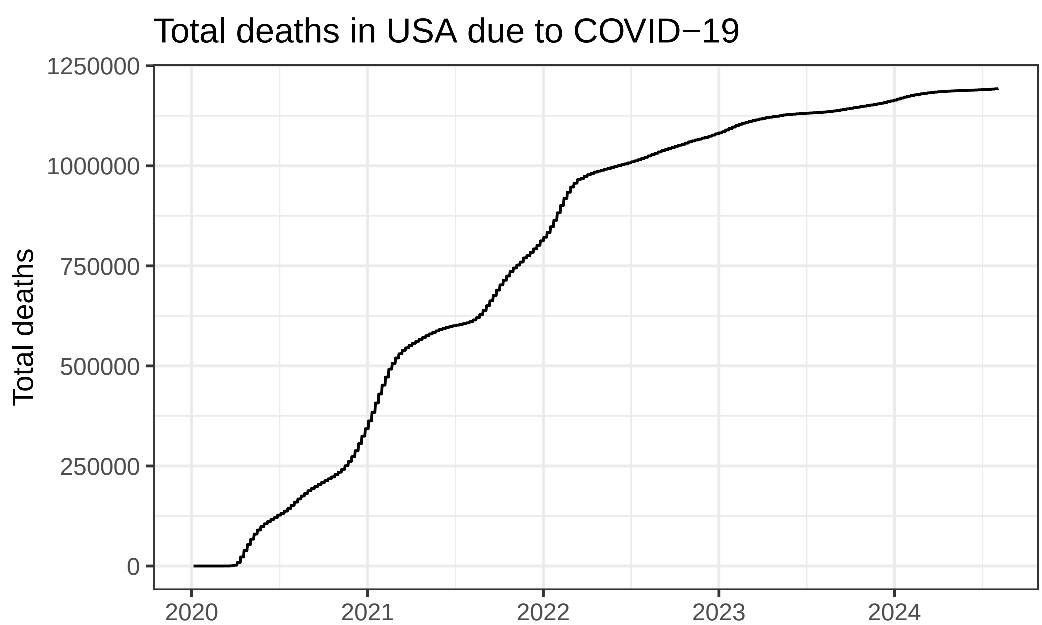
Figure: U.S. Quarterly GDP growth rate 1948:Q1 to 2016:Q1



(b) Epidemiology: the number of influenza cases observed over sometime period, the number of COVID-19 cases and deaths, etc.

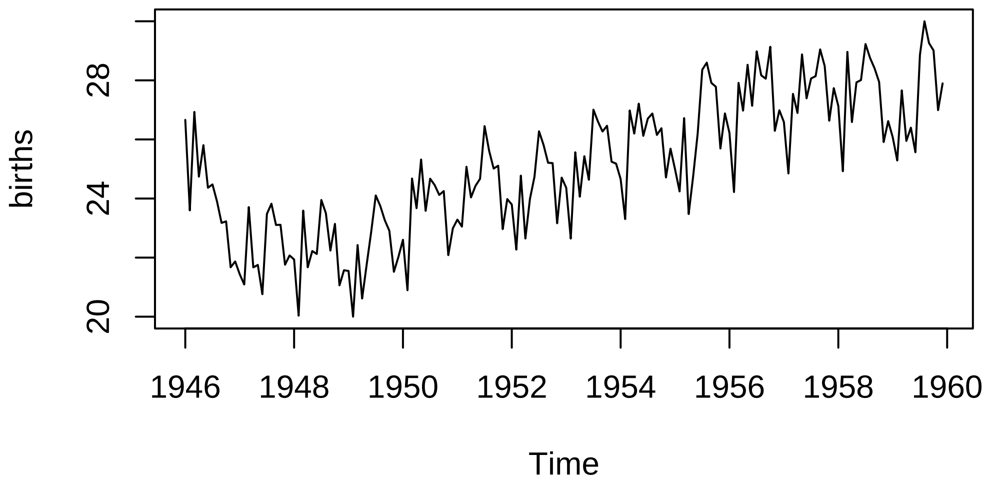
- For instance, the following figures are the time series plot of COVID-19 related death rate in Norway and USA





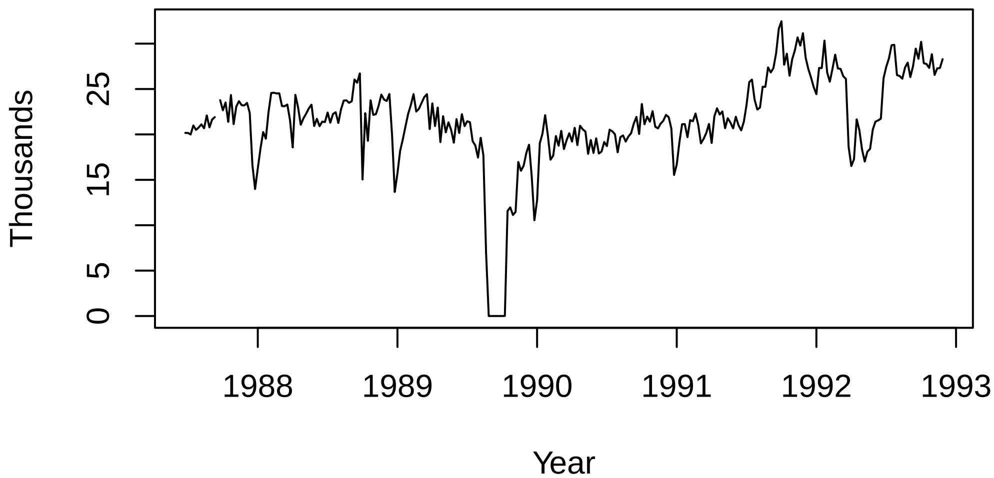
(c) Social sciences: population series, such as birthrates, school enrollments, etc.

- For example, the following is the time series plot of the monthly number of births in New York City (thousands), 1946–1959.



(d) Transport: Number of traffic accidents per day, month, or year, number of passengers traveling through airports each month or year, etc

- For example the following plot is the time series plot of the weekly economy class passengers on ansett airlines between Australia's two largest cities: Melbourne-Sydney



(e). Global warming, $CO2$ concentration in the atmosphere over time, etc.

Key features of time series data

- ordered observations.
- dependecy on time,
  - temporal dependence of a time series can be considered the main characteristic of timeseries data, which means that an observation collected at a particular time is related or influenced by the previously collected observation.
  - For instance, Norway's current monthly unemployment rate is related to the rate from the previous month.

What is Time Series Analysis/ Time Series Econometrics?

- Time series analysis (or time series econometrics) is the study of time series data.

Why study Time Series Econometrics?

- Understand patterns over time (trends, seasonality, cycles).
- Make forecasts and predictions.
- Model relationships in dynamic systems.
- Applications in economics, finance, health, climate science, etc.

Key components of time series data

Trend.

- it represent the average change (change in the mean) in the time series over time.
- it tells the overall direction of the data such as increasing, decreasing, or constant.
- When a time series represents a direction of growth, shrink and stability directions or pattern in the long term, we can consider this as the trend component of time series.

Examples of trends are:

- $T_t = \beta_0$ (constant over time , we usually refer to this case as "no trend"),
- $T_t = \beta_0 + \beta_1 t$ (linear increase or decrease over time),
- $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$ (quadratic over time), and so on.

Seasonality

- represents regular periodic fluctuations in the time series.
- regular patterns at fixed intervals (e.g., monthly sales spikes).
- reapeating patern of the data over a set period of time

Irregular/Random variations

- unpredictable, short-term fluctuations.
- is the residual or error and represents the remaining unexplained variation in the data.
- in other words, it is a random or stochastic component.

Time Series decomposition

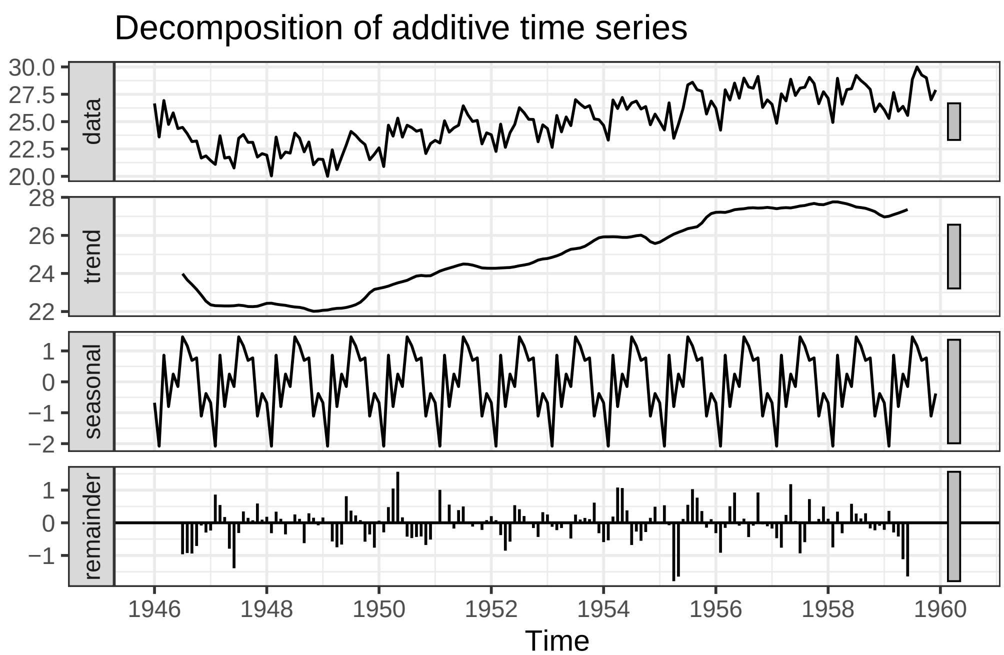
- A time series Y_t can generally be expressed as a sum of trend, seasonality and random variation

$$Y_t = T_t + S_t + \epsilon_t$$

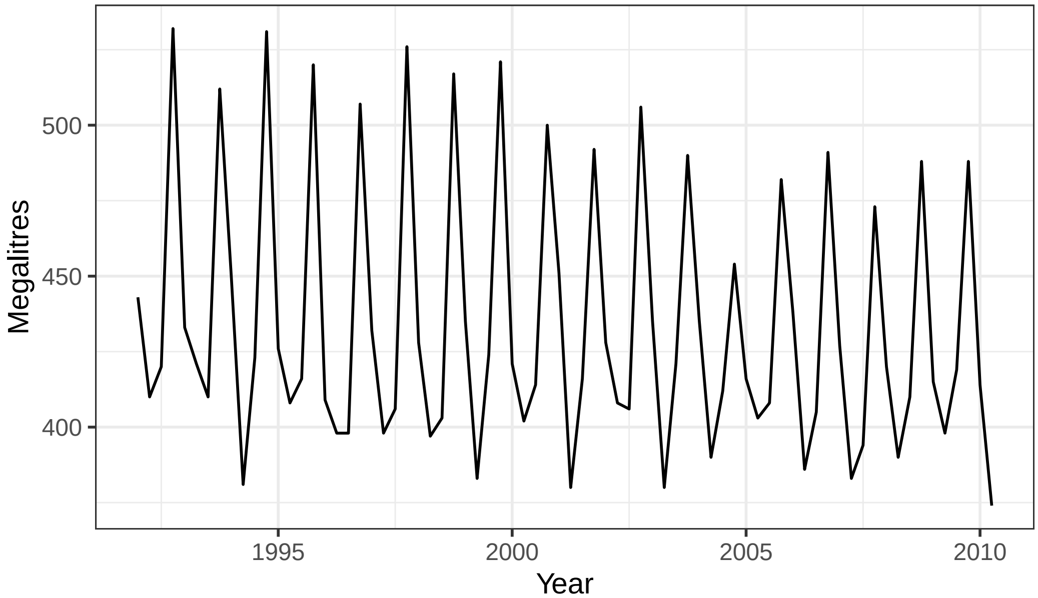
where T_t is the trend, S_t represent seasonality,and ϵ_t represents the irregular/random noise

Some examples

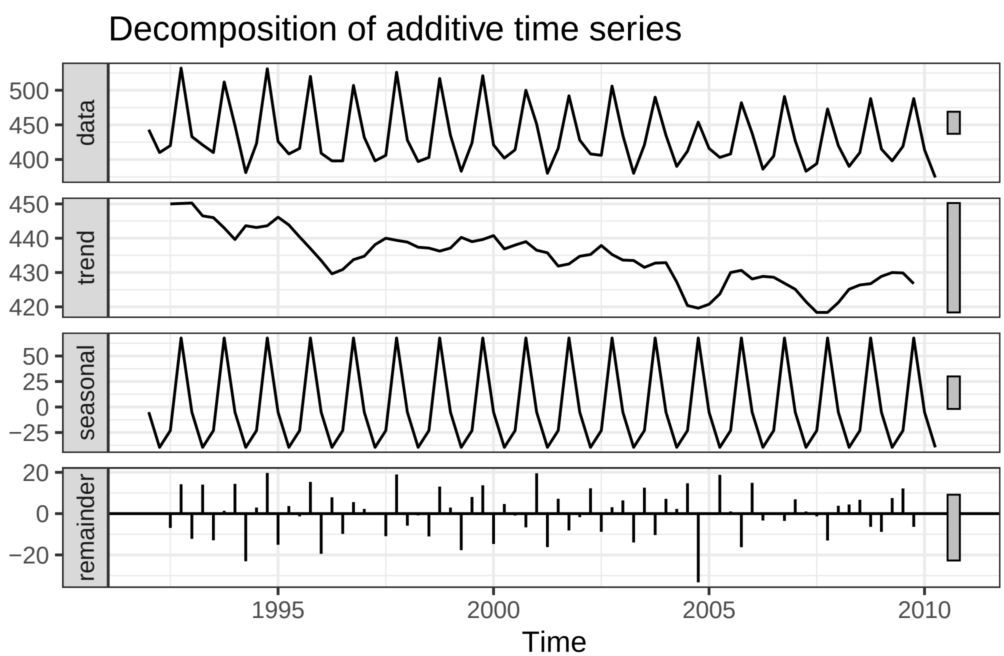
Example 1. The figure below illustrates time-series decomposition of the monthly number of births in New York City (thousands), 1946–1959.



Example 2: The time seires plot below shows the quarterly beer production in Australia



The time series decomposition of the quarterly beer production in Australia:



We want to forecast the value of future beer production. We can model this data using a regression model with a linear trend and quarterly dummy variables,

$$y_t = \beta_0 + \beta_1 t + \beta d_{2,t} + \beta d_{3,t} + \beta d_{4,t} + e_t$$

where $d_{i,t} = 1$ if t is quarter i and 0 otherwise. The first quarter variable has been omitted, so the coefficients associated with the other quarters are measures of the difference between those quarters and the first quarter.

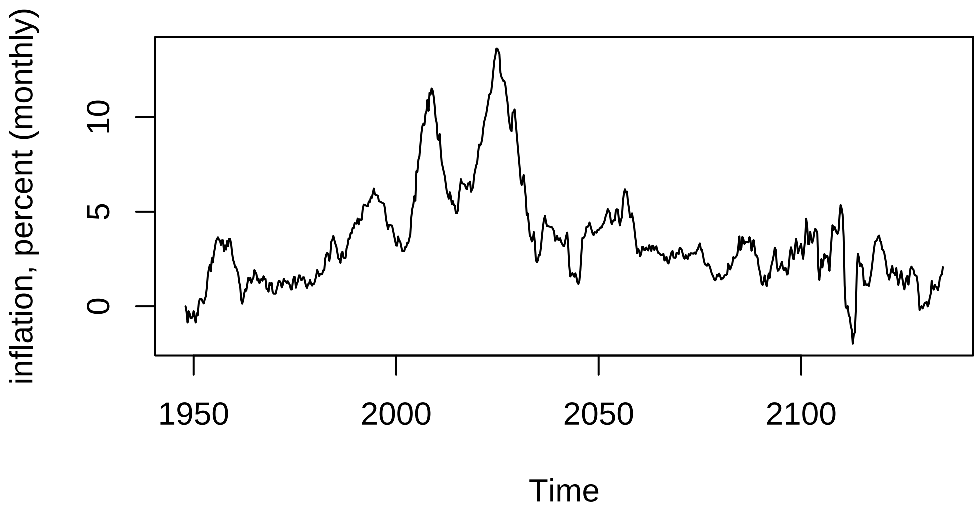
```
Call:
tslm(formula = beer2 ~ trend + season)

Residuals:
    Min      1Q  Median      3Q     Max 
-42.903  -7.599  -0.459   7.991  21.789 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 441.80044    3.73353 118.333  < 2e-16 ***
trend        -0.34027    0.06657  -5.111 2.73e-06 ***
season2     -34.65973    3.96832  -8.734 9.10e-13 ***
season3     -17.82164    4.02249  -4.430 3.45e-05 ***
season4      72.79641    4.02305  18.095  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.23 on 69 degrees of freedom
Multiple R-squared:  0.9243,    Adjusted R-squared:  0.9199 
F-statistic: 210.7 on 4 and 69 DF,  p-value: < 2.2e-16
```

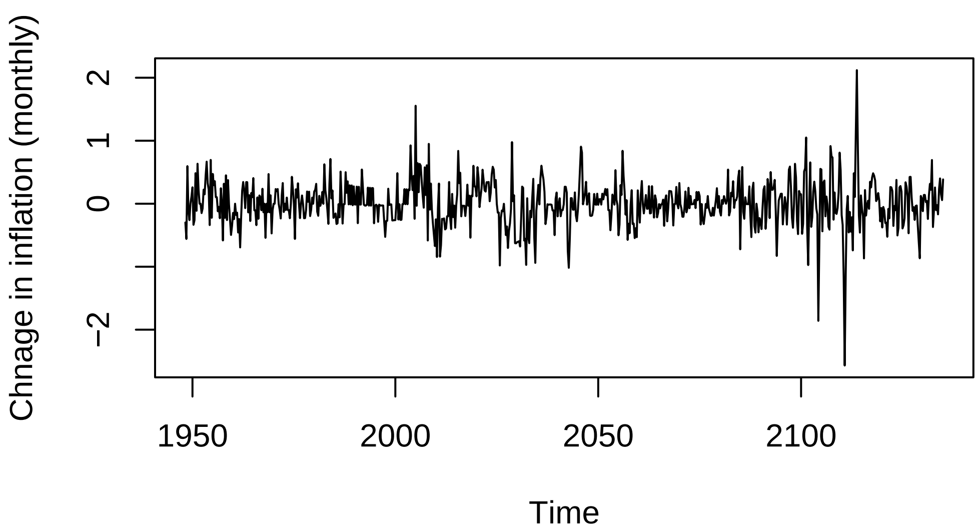
- Note that trend and season are not objects in the R workspace; they are created automatically by **tslm()** when specified in this way. Otherwise, can be created mannually.
- There is an average downward trend of -0.34 megalitres per quarter.
- On average, the second quarter has production of 34.7 megalitres lower than the first quarter, the third quarter has production of 17.8 megalitres lower than the first quarter, and the fourth quarter has production of 72.8 megalitres higher than the first quarter.

Stationary vs non-stationary time series data

Defination: - A Stationary time series data has constant means, variance and autocovariance overtime.

- Non-stationary timeseries data has means and variances that change over time
- The first figure below is non-stationary while the second one is stationary





Time Series Models

Suppose we have a time series data, Y_t.

Autoregressive(AR) models: $Y_t = \phi_1 Y_{t-1} + \epsilon_t$

Moving Average(MA) models: $Y_t = \phi_1 \epsilon_t + \epsilon_{t-1}$

ARMA models: combine AR and MA.

ARIMA models: includes differencing to address non-stationarity.

Advanced topics

- Vector Autoregression (VAR) (chapter 13)
- Cointegration and Error Correction Models (ECM) (chapter 12 and 13)
- Volatility models (e.g.,ARCH, GARCH) (chapter 14)

Summary and Resources

Key Take aways:

- Time series data is time-ordered and exhibits unique challenges.
- Understanding the characteristics of a time series—such as trends, seasonality, stationarity, and non-stationarity—is essential for effective analysis.
- the characterstics of a time series influence the choice of models and methods, ensuring accurate interpretation and forecasting.

Resources:

- **Textbook:** Principles of Econometrics 5th Edition (POE5), Wiley - 2018.By R. Carter Hill, William E. Griffiths and Guay C. Lim.
- Some useful links:
  - https://vlyubchich.github.io/tsar/l02_tsintro.html
  - https://otexts.com/fpp2/regression-intro.html
  - https://otexts.com/fpp2/useful-predictors.html
  - https://otexts.com/fpp2/expsmooth.html

Lags of time series data and regression

- **Lag:** is when you're looking at past data points to understand the current or future situation.
- In simple terms, a "lag" is like looking back in time.
- For example, if you're studying sales this month, you might look at sales from last month to predict current trends.
- So, the "lag" is the previous month's sales data compared to this month.

```
'data.frame':   273 obs. of  4 variables:
 $ dateid01: Date, format: "1948-01-01" "1948-04-01" ...
 $ g       : num  2.267 2.517 2.418 0.429 -1.888 ...
 $ inf     : num  2.119 1.592 1.684 -0.918 -0.951 ...
 $ u       : num  3.7 3.7 3.8 3.8 4.7 5.9 6.7 7 6.4 5.6 ...
 - attr(*, "var.labels")= chr [1:4] "" "U.S. growth rate in GDP" "U.S. inflation rate" "U.S u
```

```
    dateid01      g    inf   u
1 1948-01-01  2.267  2.119 3.7
2 1948-04-01  2.517  1.592 3.7
3 1948-07-01  2.418  1.684 3.8
4 1948-10-01  0.429 -0.918 3.8
5 1949-01-01 -1.888 -0.951 4.7
6 1949-04-01 -1.344 -0.109 5.9
```

Figure 1: U.S. Quarterly GDP growth rate 1948:Q1 to 2016:Q1

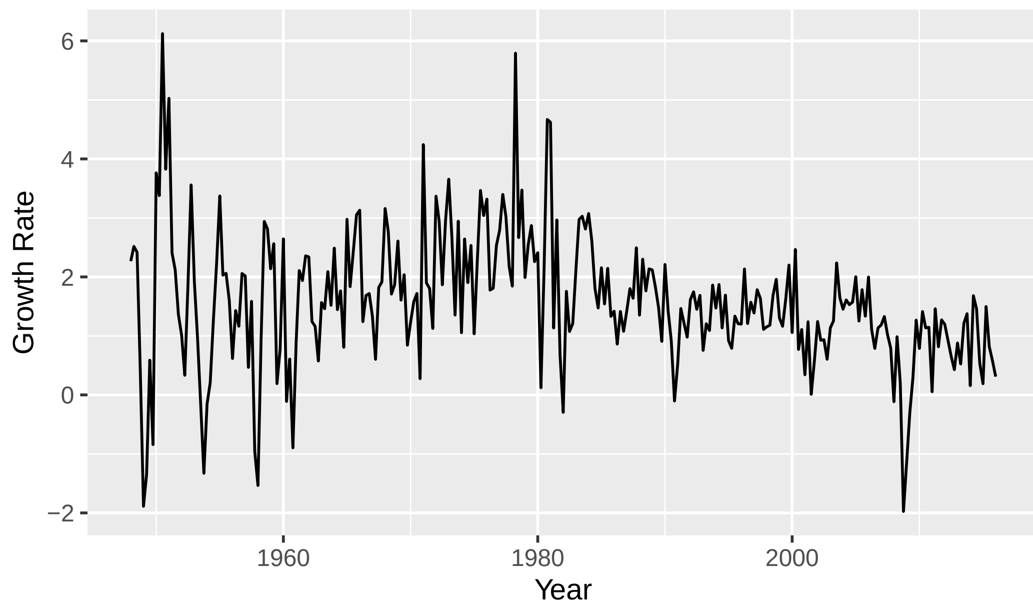
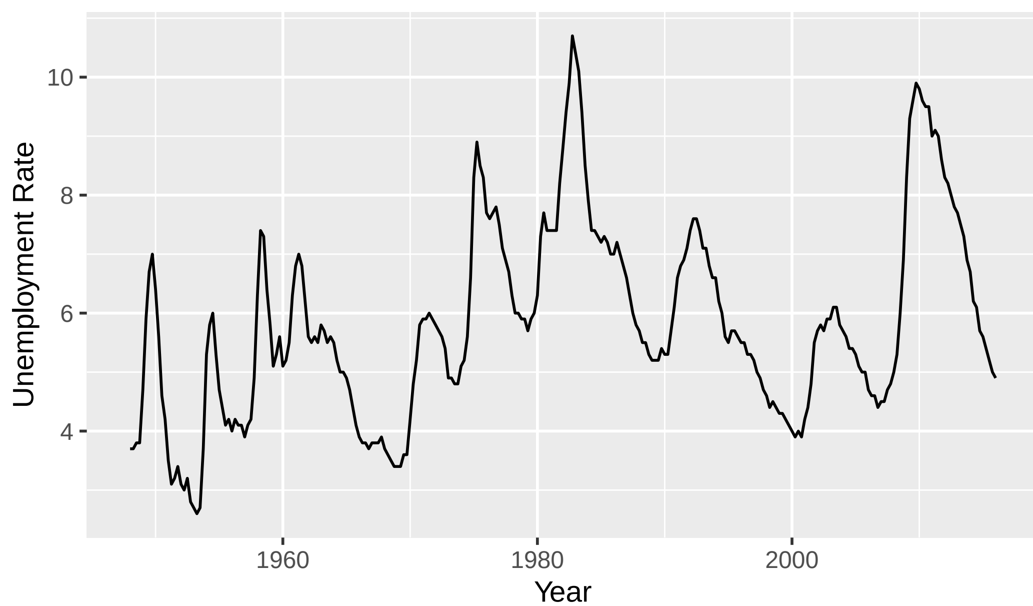


Figure 2: U.S. Quarterly unemployment rate 1948:Q1 to 2016:Q




```
    dateid01   u
1 1948-01-01 3.7
2 1948-04-01 3.7
```

```
3 1948-07-01 3.8
4 1948-10-01 3.8
5 1949-01-01 4.7
6 1949-04-01 5.9
```

```
    dateid01   u  u1  u2
1 1948-01-01 3.7  NA  NA
2 1948-04-01 3.7 3.7  NA
3 1948-07-01 3.8 3.7 3.7
4 1948-10-01 3.8 3.8 3.7
5 1949-01-01 4.7 3.8 3.8
6 1949-04-01 5.9 4.7 3.8
```

Drop rows with **NA's** values from the data frame

```
    dateid01   u  u1  u2
3 1948-07-01 3.8 3.7 3.7
4 1948-10-01 3.8 3.8 3.7
5 1949-01-01 4.7 3.8 3.8
6 1949-04-01 5.9 4.7 3.8
7 1949-07-01 6.7 5.9 4.7
8 1949-10-01 7.0 6.7 5.9
```

```
Call:
lm(formula = u ~ u1 + u2, data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.82533 -0.16751 -0.02107  0.15891  1.07456 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.28852    0.06661   4.332 2.09e-05 ***
u1           1.61282    0.04570  35.295  < 2e-16 ***
u2          -0.66209    0.04557 -14.528  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2947 on 268 degrees of freedom
Multiple R-squared:  0.9678,	Adjusted R-squared:  0.9675 
F-statistic:  4022 on 2 and 268 DF,  p-value: < 2.2e-16
```

Alternatively, more eaily using **dynlm()** function instead of **lm()** function:

```
Time series regression with "ts" data:
Start = 3, End = 273

Call:
dynlm(formula = u ~ L(u, 1) + L(u, 2), data = ts(usmacro))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.82533 -0.16751 -0.02107  0.15891  1.07456 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.28852    0.06661   4.332 2.09e-05 ***
L(u, 1)      1.61282    0.04570  35.295  < 2e-16 ***
L(u, 2)     -0.66209    0.04557 -14.528  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2947 on 268 degrees of freedom
Multiple R-squared:  0.9678,    Adjusted R-squared:  0.9675 
F-statistic:  4022 on 2 and 268 DF,  p-value: < 2.2e-16


Time series regression with "ts" data:
Start = 3, End = 273

Call:
dynlm(formula = u ~ L(u, 1:2) + L(g, 0:1), data = ts(usmacro))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.70325 -0.15332 -0.01503  0.12731  1.06536 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.40497    0.06332   6.396 7.14e-10 ***
L(u, 1:2)1   1.52327    0.04854  31.382  < 2e-16 ***
L(u, 1:2)2  -0.55882    0.04863 -11.492  < 2e-16 ***
L(g, 0:1)0  -0.14844    0.01620  -9.160  < 2e-16 ***
```

```
L(g, 0:1)1   0.02339    0.01873   1.248    0.213
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.255 on 266 degrees of freedom
Multiple R-squared:  0.976, Adjusted R-squared:  0.9757
F-statistic:  2709 on 4 and 266 DF,  p-value: < 2.2e-16

Time Series:
Start = 1
End = 5
Frequency = 1
[1] 1 2 3 4 5

Time Series:
Start = 1
End = 6
Frequency = 1
   t stats::lag(t, -1)
1  1                NA
2  2                 1
3  3                 2
4  4                 3
5  5                 4
6 NA                 5

Time Series:
Start = 0
End = 5
Frequency = 1
   t stats::lag(t)
0 NA            1
1  1            2
2  2            3
3  3            4
4  4            5
5  5           NA

Time Series:
Start = 0
End = 6
```

```
Frequency = 1
   t stats::lag(t) stats::lag(t, -1)
0 NA             1                NA
1  1             2                NA
2  2             3                 1
3  3             4                 2
4  4             5                 3
5  5            NA                 4
6 NA            NA                 5

Time Series:
Start = 0
End = 6
Frequency = 1
   t stats::lag(t) stats::lag(t, -1)
0 NA             1                NA
1  1             2                NA
2  2             3                 1
3  3             4                 2
4  4             5                 3
5  5            NA                 4
6 NA            NA                 5

Time Series:
Start = 2
End = 4
Frequency = 1
  t stats::lag(t) stats::lag(t, -1)
2 2             3                 1
3 3             4                 2
4 4             5                 3
```

Chapter 09

- next we start chapter 09