PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Introduction to Machine Learning

MLEARN 510A – Lesson 9

Recap of Lesson 8

- Dimensionality Reduction
- Issues Encountered in High Dimensional Settings
- Principal Component Analysis (PCA)
- Principal Components Regression (PCR)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)

Outline of Lesson 9

- Introduction to Time Series
- Time Series Decomposition
- Time Series Forecasting
- ETS Models
- ARIMA Models
- Model Evaluation

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY *of* WASHINGTON

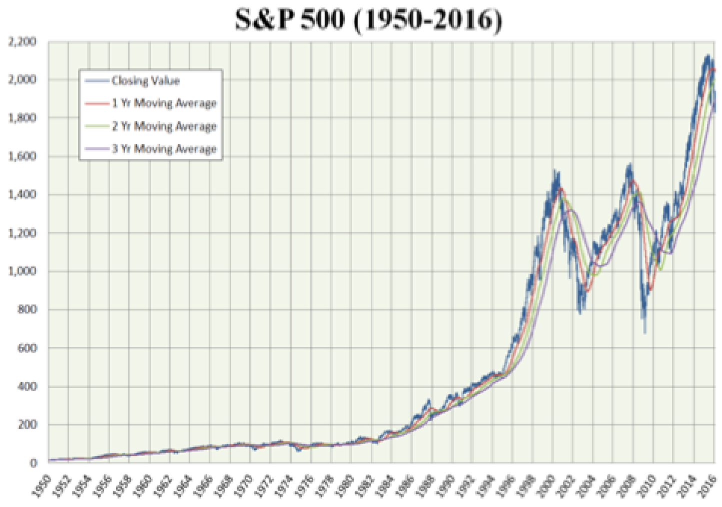
Introduction to Time Series

Time Series Data

- **Time Series:** A series of data points indexed (or listed or graphed) in an order of time

- Represents a sequence taken at successive equally spaced intervals in time

- Examples: Daily stock prices, quarterly sales figures



What Makes Time Series Different?

- **Order matters:** Changing the order could change the meaning of data
  - Note the difference with say, predicting wages based on education levels, where the individuals' data can be entered in any order

- **Correlation**: Correlation exists between data points
  - Note the difference with say, classifying species of flowers based on their attributes where individual data points are independent of each other

Challenges

- Much more challenging to work with times series compared to traditional supervised learning

- Common assumptions in standard supervised learning
  - Independent and identically distributed (i.i.d) data
  - Distributions for training and test data are similar
  - Distributions fixed over time

- *None of these assumptions hold in typical time series data!*

Time Series Analysis

- **Time Series Analysis**: comprises methods for analyzing time series data to extract meaningful statistics

- Usually divided into two categories

- **Frequency-Domain Methods**
  - Spectral Analysis, Wavelet Analysis

- **Time-Domain Methods**
  - Auto-correlation analysis, cross-correlation analysis

Stationarity

- **Strong Stationarity** (also called Strict-Sense Stationarity)
  - Distribution of a finite sub-sequence of random variables of the stochastic process remains the same as we shift it along the time index axis

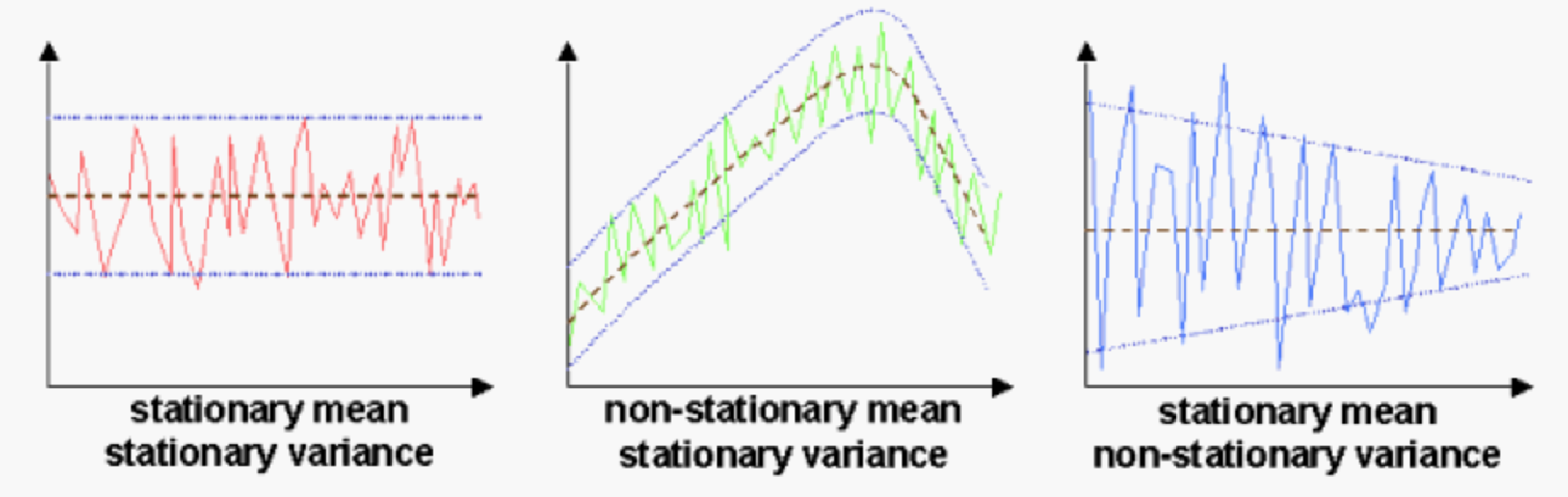
$$F_X(x_{t_{1+\tau}}, \ldots, x_{t_{n+\tau}}) = F_X(x_{t_1}, \ldots, x_{t_n})$$

- **Weak Stationarity** (Also called Wide-Sense Stationarity)
  - Requires the shift-invariance (in time) of the first moment and the cross moment (the auto-covariance)
  - This implies that the process has the same mean at all time points $E[x_i]=\mu$, and
  - Covariance between the values at any two time points, *t* and *t−k*, depend only on *k*, the difference between the two times, and not on the location of the points along the time axis

$$cov(X_{t_1}, X_{t_2}) = K_{XX}(t_1, t_2) = K_{XX}(t_2 - t_1, 0) = K_{XX}(\tau)$$

$$F_{t_1,\ldots,t_n}(x_{t_1}, \ldots, x_{t_n}) = P(X_{t_1}(\omega) \le x_{t_1}, \ldots X_{t_n}(\omega) \le x_{t_n})$$

Some Examples



https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322

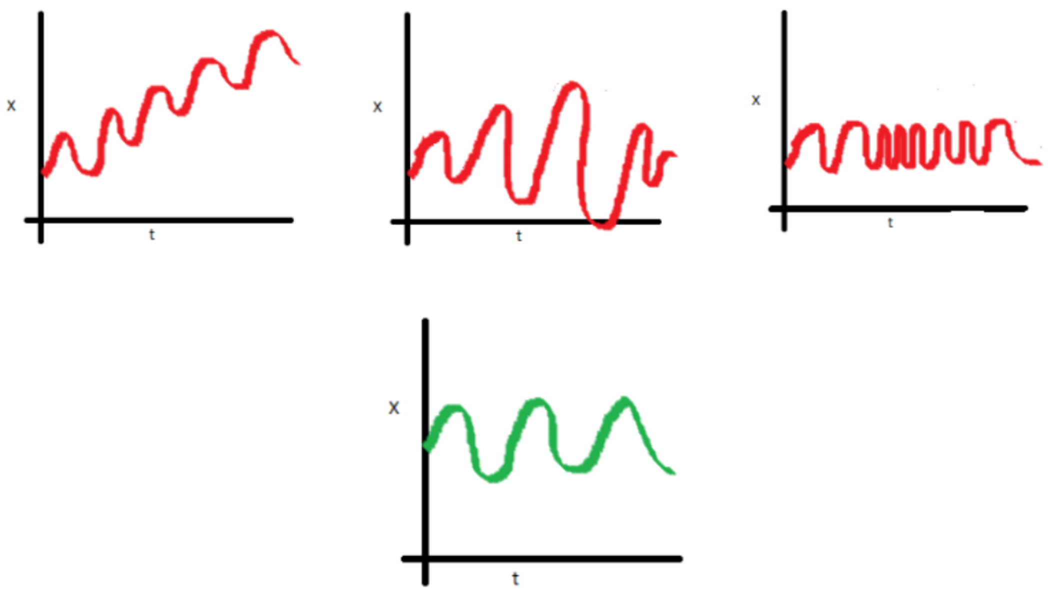
Quiz

- Can you think of any reason why we would desire to have a stationary time series?

Why is Stationarity Useful?

- Stationary processes are a sub-class of a wider family of possible models of reality. This sub-class is easier to model and investigate

- Such processes are easier to predict, as the *way* they change is predictable



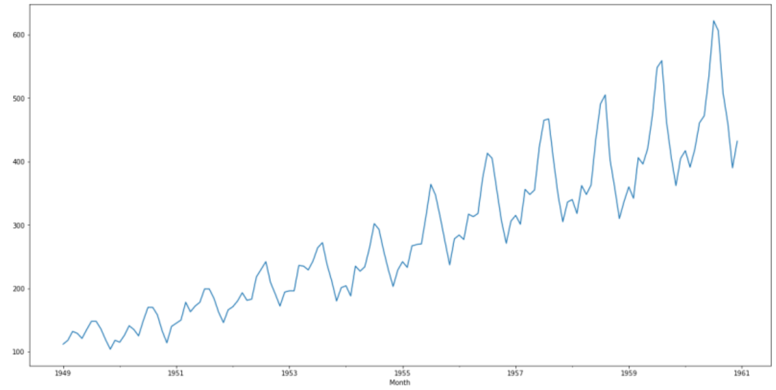
Non-Stationary Time Series

- Most time series processes are not inherently stationary
- *A time series is non-stationary if its statistical properties change over time*
- When the assumption of stationarity fails, parameters of interest may no longer be a constant
  - In this case, they are naturally modeled as functions of time, which are infinite dimensional objects
- In these cases, we can apply some processing techniques to make a time series stationary

Methods to Test Stationarity

- There are various methods test whether a time series is stationary

- **Visual**: Plot data and see if statistical properties change with time



- **Statistical Test:** Use *Augmented Dickey-Fuller* (ADF) test
  - **Null Hypothesis**: The series has a unit root (value of a =1)
  - **Alternate Hypothesis**: The series has no unit root
  - Failure to reject null hypothesis implies that the series is non-stationary

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Time Series Decomposition

Patterns in Time Series

- **Trend**: A pattern exists involving a long-term progression (increase or decrease) in the data

- **Seasonal:** A *periodic* pattern exists due to the calendar (e.g., daily, monthly, quarterly, day of the week). Seasonality is always of a fixed and known period

- **Cyclical**: A pattern exists where the data exhibits rises and falls that are not of fixed period. The duration of these fluctuations is usually of at least two years

- **Irregular (noise)**: Describes random and irregular influences

Seasons vs. Cycles?

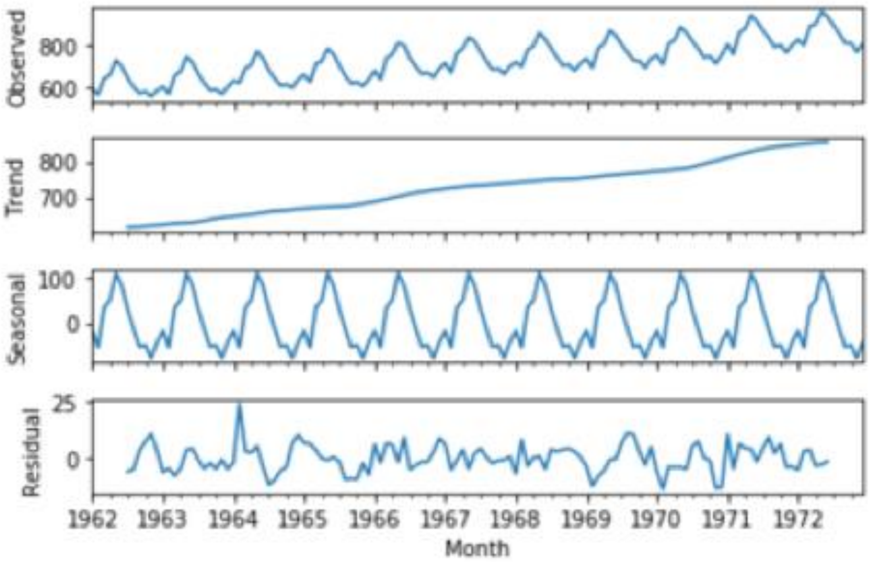
- Seasonal patterns usually have a constant length while cyclical pattern exhibit variable length
- The average length of a cyclic pattern is longer than the length of a seasonal pattern
- The magnitude of cycle shows more variations than the magnitude of seasonal pattern
  - This makes it harder to predict cyclic data compared to seasonal data

Time Series Decomposition

- Decomposition is a statistical task that deconstructs a time series into several components
  - Trend
  - Seasonality
  - Irreducible component

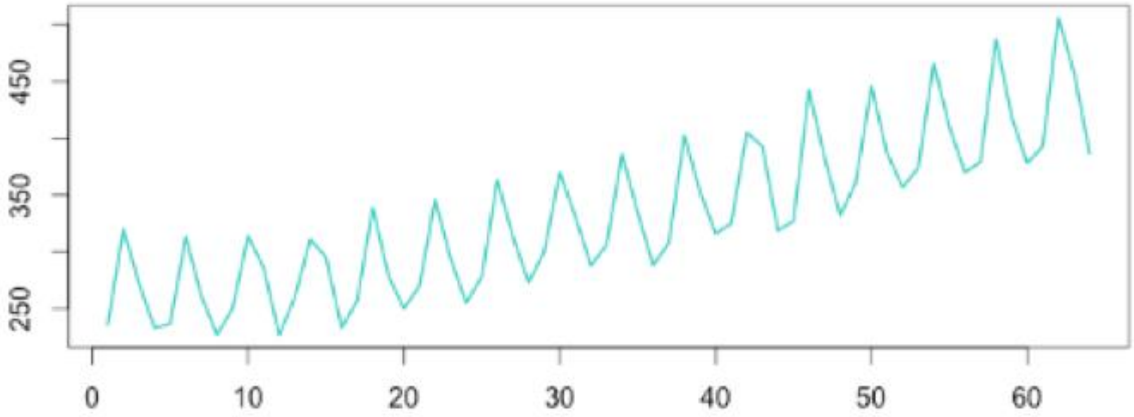
Time Series Decomposition Plot

- The first plot shows the actual time series
- The trend line indicates the general tendency of the time series
- The seasonal portion shows that there's a seasonal pattern
- The residual is the difference between the observed value and the trendline & seasonal estimates

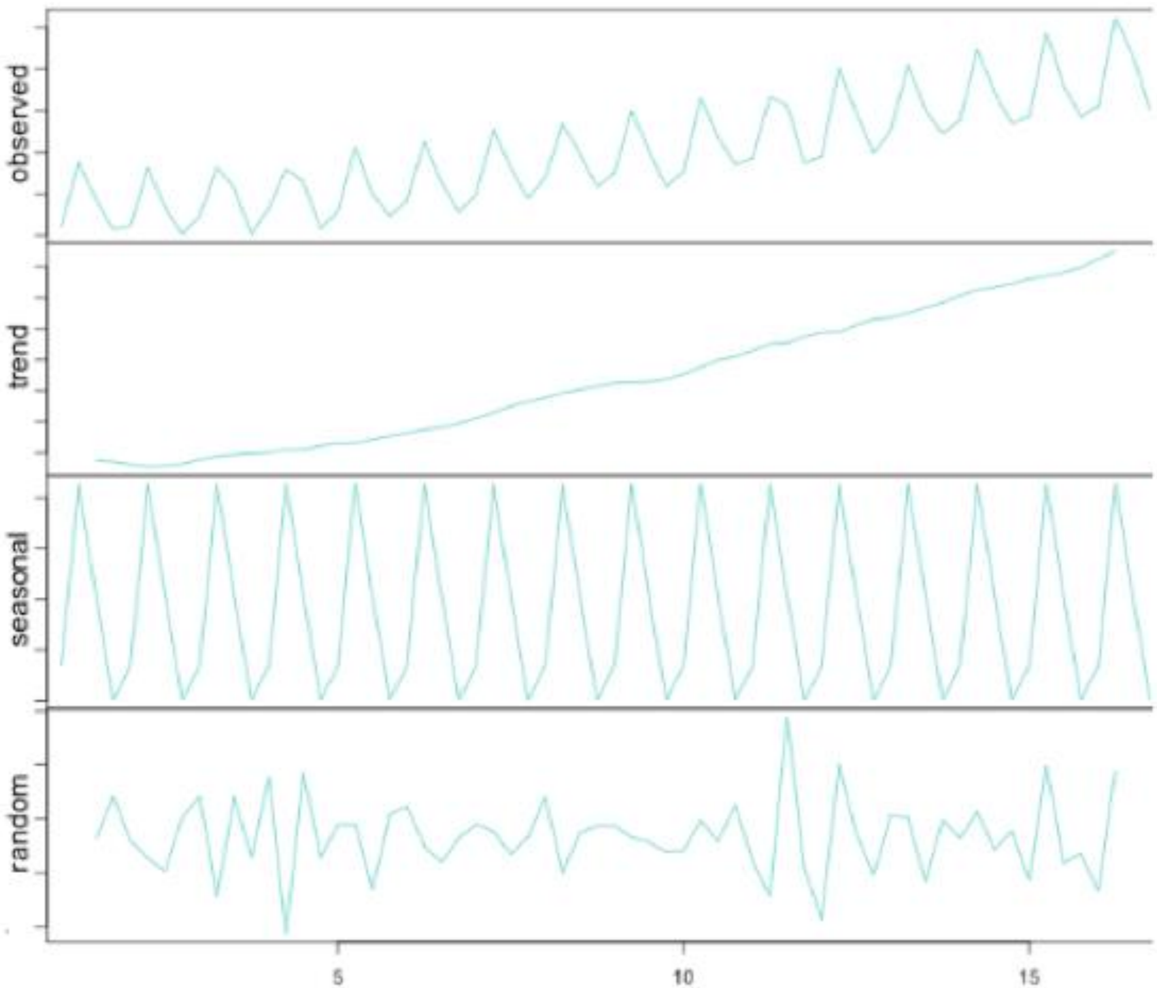


Time Series Decomposition – Additive

- Additive Model: Time series = Seasonal + Trend + Random
- Can be written as $y_t = T_t + S_t + R_t$
- The additive model is useful when the seasonal variation is relatively constant over time



Australian Beer Production – Additive Decomposition

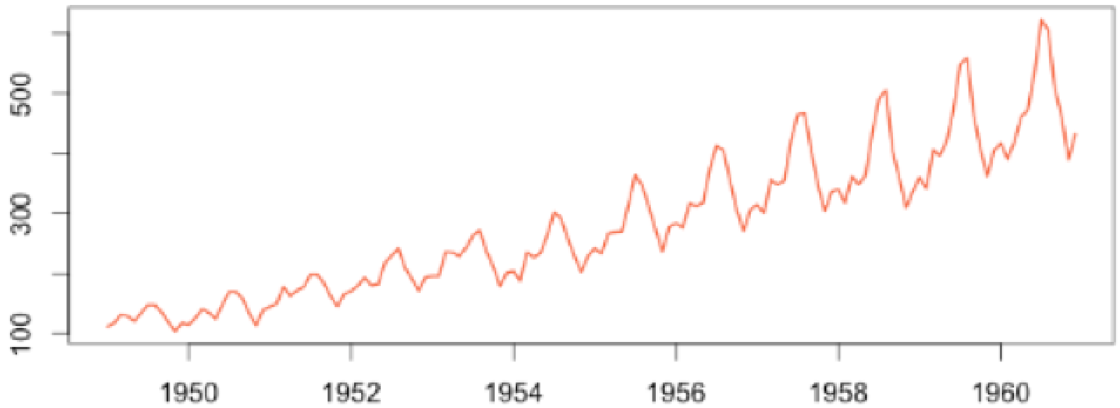


Time Series Decomposition – Multiplicative

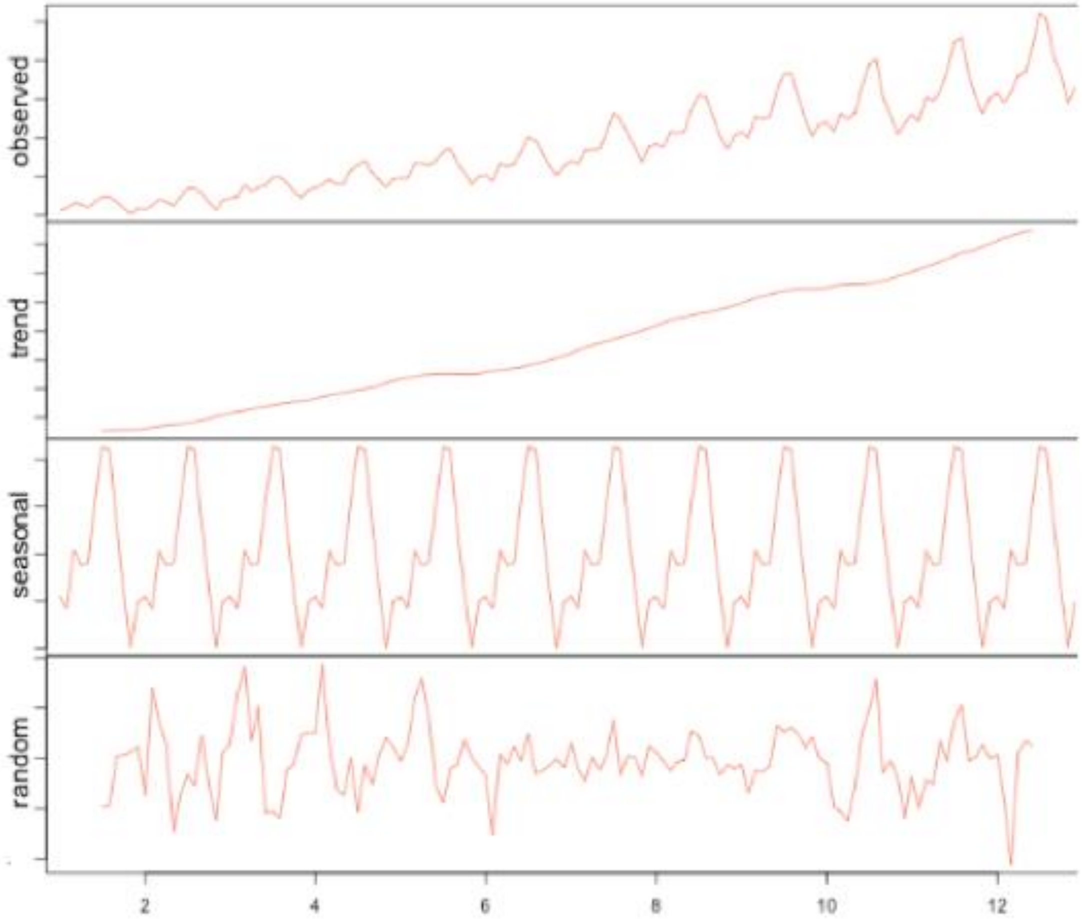
- Multiplicative Model: Time series = Seasonal x Trend x Random

- Can be written as $y_t = T_t \times S_t \times R_t$

- The multiplicative model is useful when the seasonal variation increases over time



Airline Passenger Numbers – Multiplicative Decomposition



Steps to Decompose Time Series

- **Step 1**: Estimate the trend
  - Use smoothing procedure such as moving averages (more later)
  - Another approach is to model the trend with a regression equation

- **Step 2**: “De-trend” the series
  - Additive decomposition: Subtract trend estimates from series
  - Multiplicative decomposition: Divide the series by trend estimates

- **Step 3**: Estimate seasonal factors using the de-trended series
  - Average the de-trended values for a specific season
  - For instance, to get a seasonal effect for January, we average the de-trended values for all Januarys in the series, and so on

- **Step 4**: Determine the random (irregular) component
  - Additive model: random = series – trend – seasonal
  - Multiplicative model: random = series / (trend*seasonal)

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY *of* WASHINGTON

Time Series Forecasting

Time Series Forecasting

- *“It’s Difficult to Make Predictions, Especially About the Future”*
  - *- Attributed to Niels Bohr*

- **Time series *forecasting*** is the use of a model to predict future values based on previously observed values

- *One-step forecast*: predicting the observation at the next time step

- *Multi-step forecast*: predicting the observation for the next *K* steps

Quiz

- Which of the following business situations require time series forecasting (select all that apply):

  1- Monthly Seattle bike rentals

  2- A stock’s daily closing value

  3- Amount of time to close a sales cycle

  4- Retail trade area analysis and demographic profiling

  5- Annual Orca population

Simple Forecasting Methods

- **Average method**: Forecasts of all future values are equal to the mean of historical data

- **Naïve method**: Simply set forecasts equal to last observed value

- **Seasonal naïve method**: Set forecasts equal to the last observed value from the same season of the year (e.g., the same month of the previous year)

- **Drift method**: Variation on the naïve method to allow forecasts to increase or decrease over time
  - The amount of change over time (called the **drift**) is set to be the avera change seen in the historical data

Common Approaches to Forecasting

- **ETS Model**
  - Simple Exponential Smoothing (SES)
  - Holt Winter’s Exponential Smoothing (HWES)
  - Holt-Winters Seasonal Method

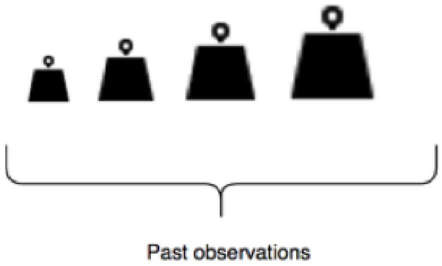
- **Autoregressive Moving Average**
  - Autoregression (AR)
  - Moving Average (MA)
  - Autoregressive Moving Average (ARMA)
  - Autoregressive Integrated Moving Average (ARIMA)
  - Seasonal Autoregressive Integrated Moving-Average (SARIMA)

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

ETS Models

Introduction to ETS Models

- Exponential smoothing forecasts use weighed averages of past observations, giving more weight to the most recent observation



- The ETS term represents how **Error**, **Trend** and **Seasonality** are applied in the smoothing method calculation

- This label can also be thought of as **E**xponen**T**ial **S**moothing

Simple Exponential Smoothing (SES)

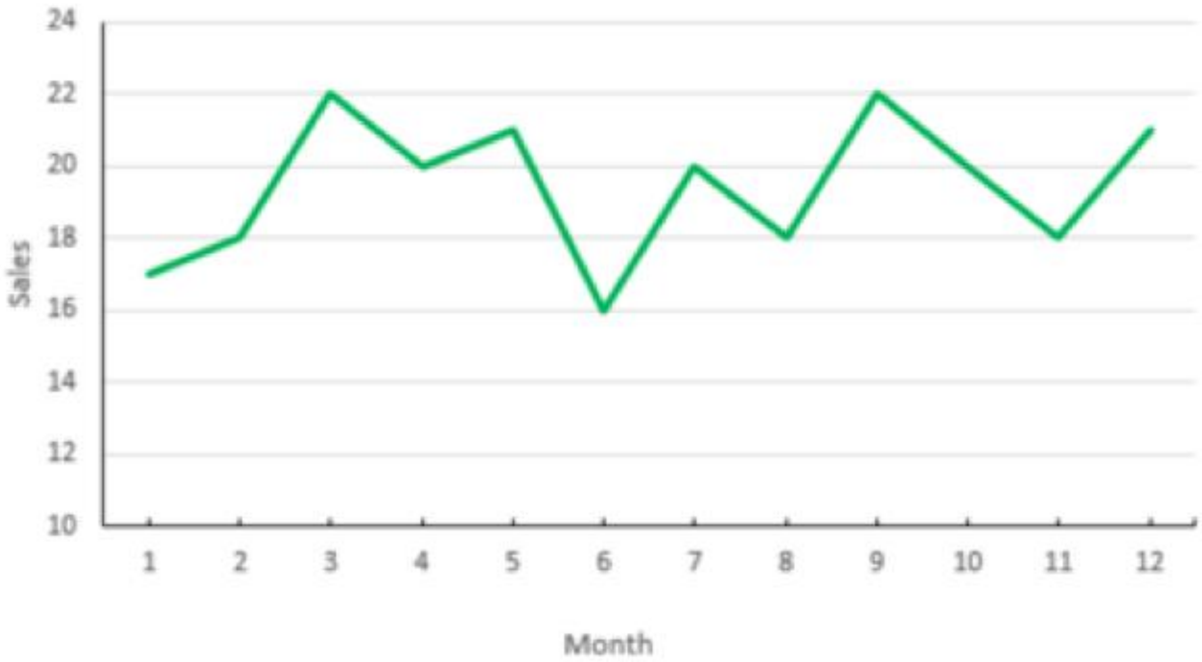
- Forecasts are calculated using weighted averages, where the weights decrease exponentially for observations further in the past
- The smallest weights are associated with the oldest observations

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha) y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots$$

- For any smoothing parameter α between 0 and 1, the weights attached to the observations decrease exponentially as we go back in time, hence the name “exponential smoothing”
  - Small α (i.e., close to 0) → more weight given to observations from the more distant past
  - Large α (i.e., close to 1) → more weight given to recent observations
  - α=1 → forecasts are equal to the naïve forecasts

Simple Exponential Smoothing

➢ Suitable for forecasting data with no clear trend or seasonal pattern



Alternative Representation – Component Form

- An alternative representation is the component form. For SES, the only component included is the level, ℓ_t

- Comprises a forecast equation and a smoothing equation for each of the components included in the method

$$\begin{aligned} \text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t \\ \text{Smoothing equation} \quad & \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \end{aligned}$$

- ℓ_t is the level (or the smoothed value) of the series at time t. Setting h=1 gives the fitted values, while setting t=T gives the true forecasts beyond the training data

Holt’s Linear Trend Method

- Holt (1957) extended simple exponential smoothing by including not only the level but also **the trend** to allow the forecasting of data
- This method involves a forecast equation and two smoothing equations (one for the level and one for the trend)
- Holt’s linear method is a very useful model to apply to any non-seasonal data set

Holt’s Linear Trend Method

$$
\begin{aligned}
\text{Forecast equation} \quad & \hat{y}_{t+h|t} = \ell_t + hb_t \\
\text{Level equation} \quad & \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
\text{Trend equation} \quad & b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}
\end{aligned}
$$

- ℓ_t denotes an estimate of the level of the series at time t
- b_t denotes an estimate of the trend (slope) of the series at time t
- α is the smoothing parameter for the level, $0 \le \alpha \le 1$
- β^* is the smoothing parameter for the trend, $0 \le \beta^* \le 1$

Damped Trend Methods

- Holt's Linear Trend is good for forecasting a time series without seasonality

- However, this method displays a constant trend (increasing or decreasing) indefinitely into the future

- Empirical evidence indicates that these methods tend to over-forecast, especially for longer forecast horizons

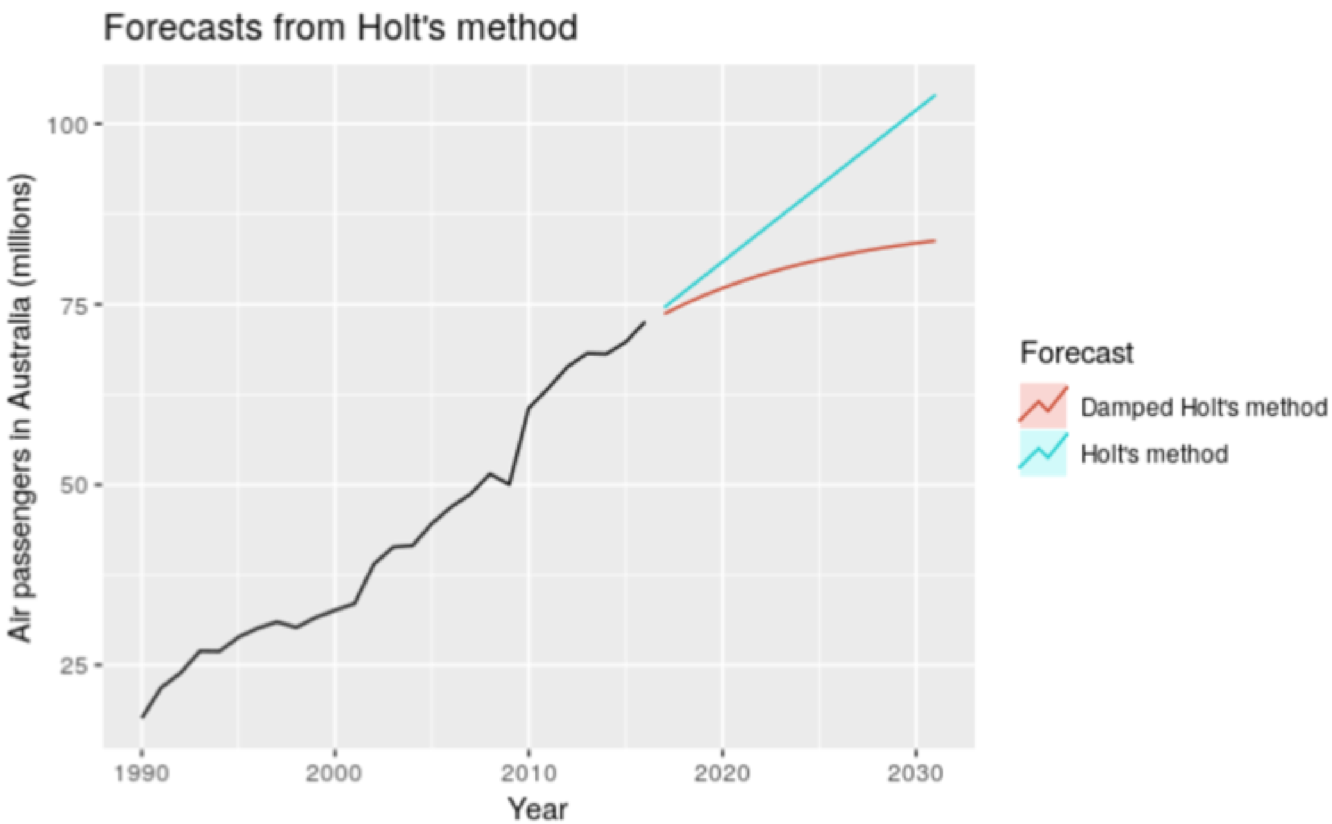
- **Damped Trend Methods:** Include a parameter that dampens the trend line into a flat line, some time into the future

Damped Trend Methods

$$\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^h)b_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$$

- ϕ=1 → method is identical to Holt's linear method
- For values between 0 and 1, ϕ dampens the trend so that it approaches a constant some time in the future
- Short-run forecasts are trended while long-run forecasts are constant
- In practice, ϕ is rarely less than 0.8 as the damping has a very strong effect for smaller values

Holt’s Method vs. Damped Trend Method



Holt-Winters Seasonal Method

- Holt (1957) and Winters (1960) extended Holt's method to capture seasonality
- Comprises of forecast equation with three smoothing equations, one for the level, one for the trend, and one for the seasonal component

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\
\ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},
\end{aligned}
$$

- where k is the integer part of (h−1)/m, which ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample

ETS Models So Far …

- **Simple Exponential Smoothing**
  - Finds the level of the time series
- **Holt’s Linear Trend**
  - Finds the level of the time series
  - Additive model for linear trend
- **Holt-Winters Seasonal**
  - Finds the level of the time series
  - Additive for trend
  - Multiplicative and Additive for seasonal components

ETS (Error, Trend, Seasonality)

- All the models we have studied thus far can be generalized using a naming system for ETS

- For each component in the ETS system, we can assign None, Multiplicative, or Additive (N, M, A) for trend and seasonality components in our time series

Table 8.5: A two-way classification of exponential smoothing methods.

Trend Component	**Seasonal Component**		
	N	**A**	**M**
	(None)	**(Additive)**	**(Multiplicative)**
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)

ETS (Error, Trend, Seasonality)

- Examples:
- A time series model that has a linear trend, and increasing seasonal components implies **ETS(A,M)**
- A time series model that has exponential trend, and no seasonality implies **ETS(M,N)**

Quiz

Short hand	Method
(N,N)	Simple exponential smoothing
[redacted]	Holt's linear method
(A_d,N)	Additive damped trend method
[redacted]	Additive Holt-Winters' method
(A,M)	Multiplicative Holt-Winters' method
[redacted]	Holt-Winters' damped method

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Auto Regressive Integrated Moving Average (ARIMA) Models

Autocorrelation in Time Series

- Autocorrelation refers to the degree of correlation between the values of the same variables across different observations

- In other words, autocorrelation measures the linear relationship between *lagged values* of a time series

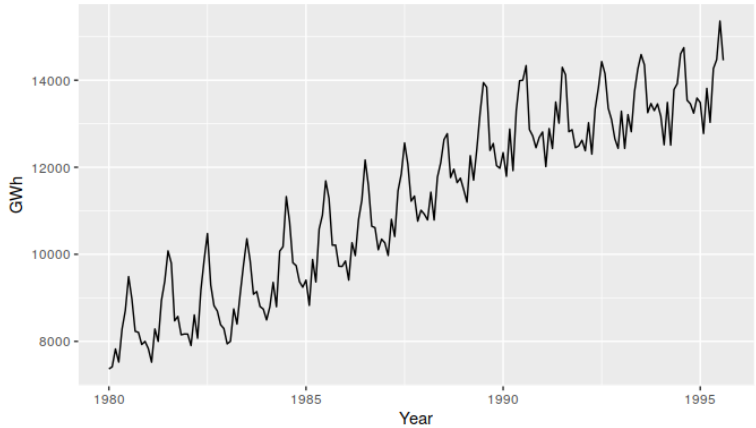
- For example, one might expect the air temperature on the 1st day of the month to be more similar to the temperature on the 2nd day compared to the 31st day

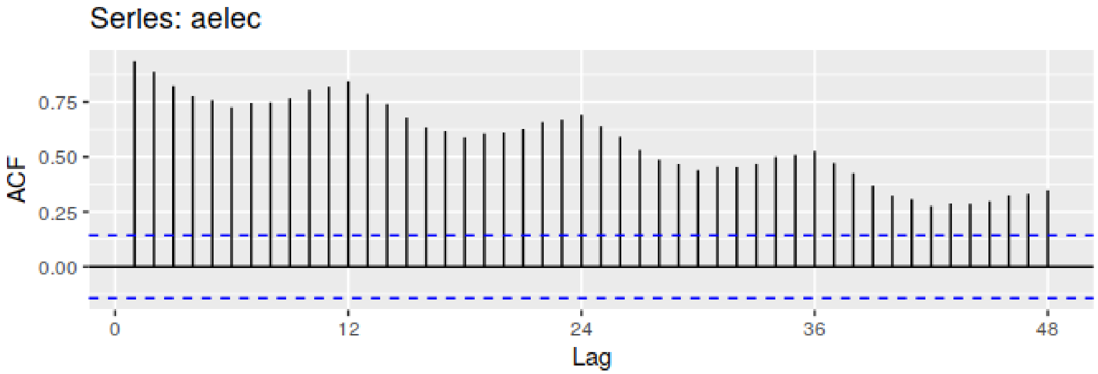
- There are several autocorrelation coefficients, corresponding to each panel in the lag plot.
  - For example, r_1 measures the relationship between y_t and y_{t-1}, r_2 measu[illegible] relationship between y_t and y_{t-2}, and so on

Autocorrelation Function Plot (ACF)

- A plot of the autocorrelation of a time series by lag is called the **A**uto**C**orrelation **F**unction, or the acronym ACF
  - This plot is also sometimes called a *correlogram* or an *autocorrelation plot*

- When data have a trend, the autocorrelations for small lags tend to be large and positive. When data are seasonal, the autocorrelations will be larger at the seasonal lags. When data are trended and seasonal, you see a combination of these effects





Quiz

Below are four time series plots and four autocorrelation plots. Try to understand which ACF plot corresponds to which series and explain how you would build a model for each one (which transformations to use etc.).

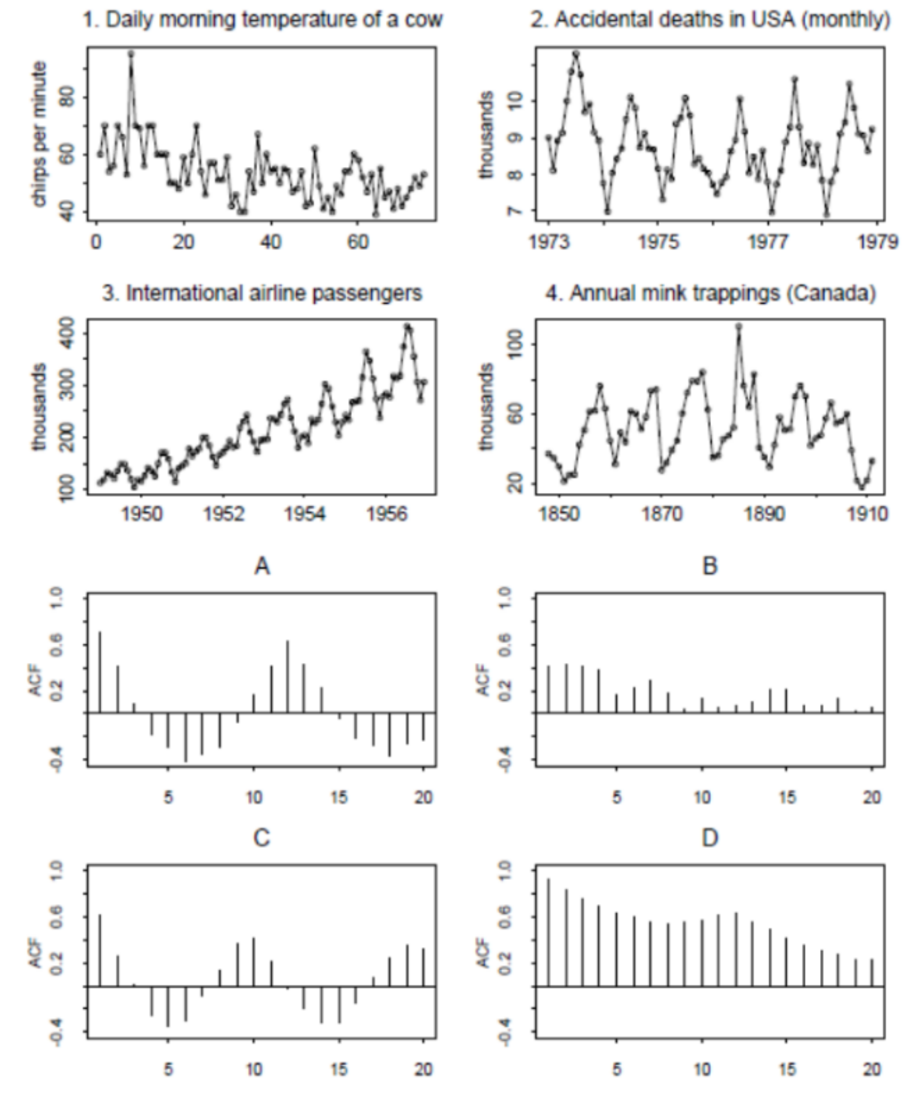
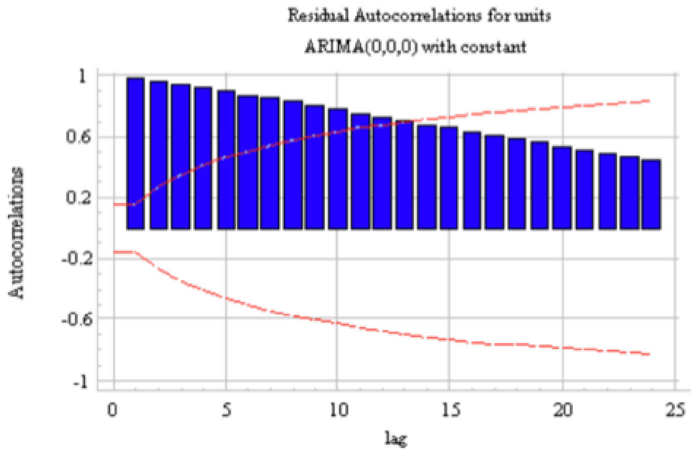
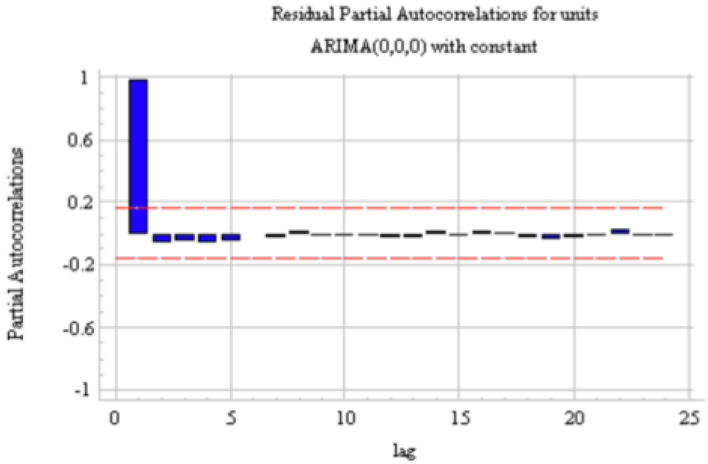


Figure 2-13: *Four time series with their ACFs.*

Partial Autocorrelation Function (PACF)

- A partial *auto*correlation is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all *lower-order*-lags





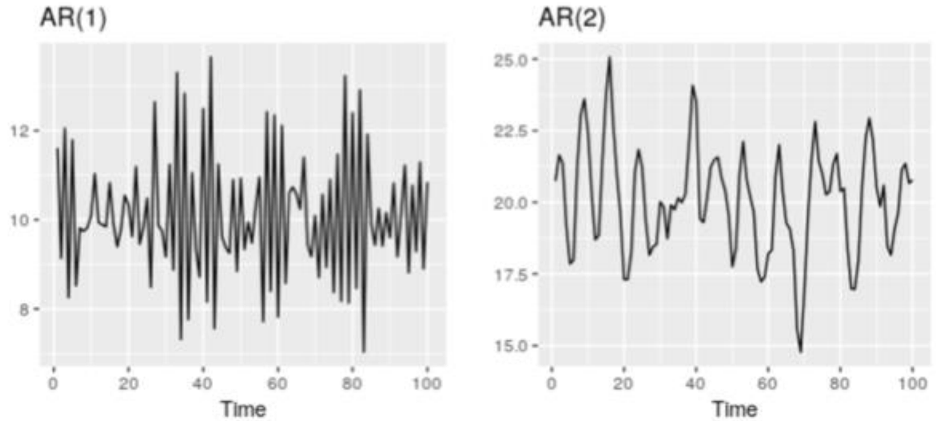
- Note that the PACF plot has a significant spike only at lag 1, meaning that all the higher-order autocorrelations are effectively explained by the lag-1 autocorrelation

Autoregressive (AR) Model

- AR model make forecasts using a linear combination of *past values of the variable*. The term *auto*regressive indicates that it is a regression of the variable against itself

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

- This is like a multiple regression but with *lagged values* of y_t as predictors. We refer to this as an **AR(*p*) model**, an autoregressive model of order *p*

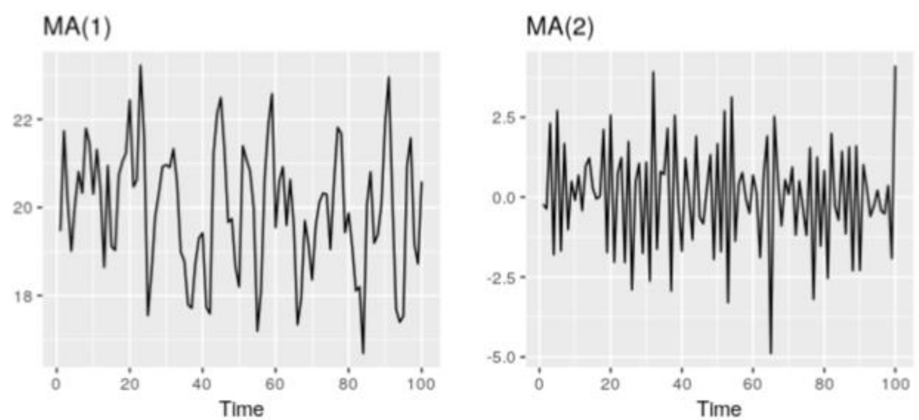


Moving Average (MA) Model

- A *Moving Average* (MA) model uses past forecast errors in a regression-like model

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

- We refer to this as an **MA(q) model**, a moving average model of order q. Since we do not *observe* the values of ε_t, so it is not really a regression in the usual sense
- Moving average *models* should not be confused with the moving average *smoothing* we discussed earlier



Autoregressive Moving Average (ARMA) Models

- The notation ARMA(*p*, *q*) refers to the model with *p* autoregressive terms and *q* moving-average terms. This model contains the AR(*p*) and MA(*q*) models

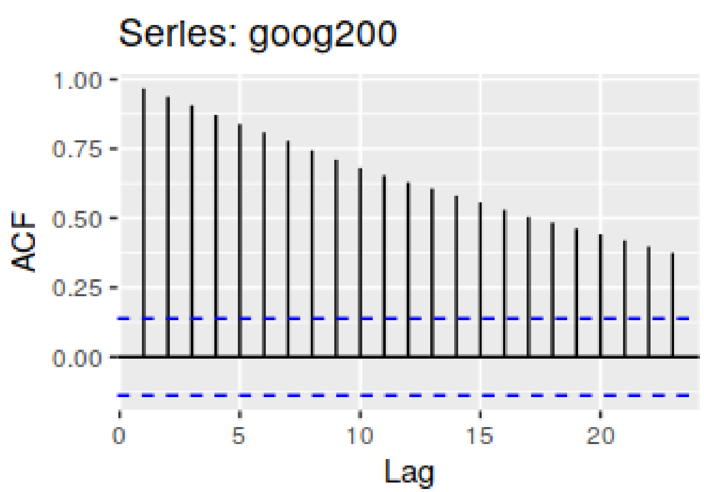
$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

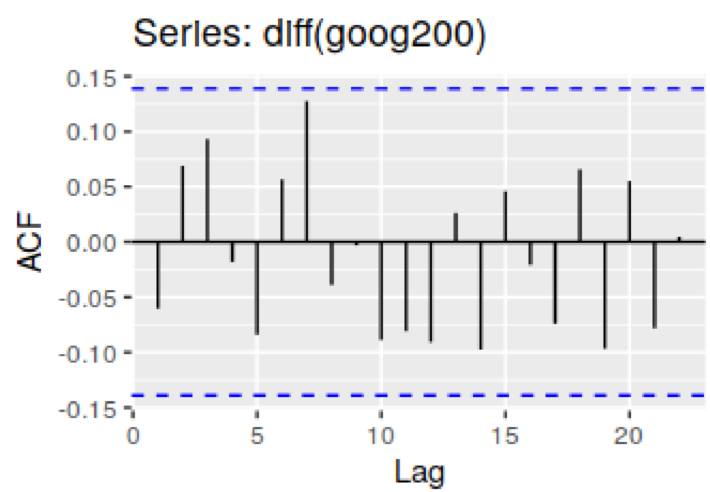
- φ = the autoregressive model's parameters
- θ = the moving average model's parameters
- c = a constant
- ε = error terms (white noise)

Differencing

- **Differencing:** One way to make a non-stationary time series stationary by computing the differences between consecutive observations

- Differencing can help stabilize the mean of a time series by removing changes in the level of a time series, and therefore eliminating (or reducing) trend and seasonality.





Different Types of Differencing

- **First Order Differencing**

$$y'_t = y_t - y_{t-1}.$$

- **Second Order Differencing**

$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

- **Seasonal Differencing**

$$y'_t = y_t - y_{t-m},$$

where $m =$ the number of seasons. These are also called "lag-m differences", as we subtract the observation after a lag of m periods.

Autoregressive Integrated Moving Average (ARIMA) Models

- Combines differencing with autoregression and a moving average model, we obtain a non-seasonal ARIMA model

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

- where y'_t is the differenced series (it may have been differenced more than once)
- *p* AR terms, *q* MA terms

ARIMA(p, d, q) Representation

- An **ARIMA(**p,d,q**)** model is composed of *p* AR terms, *q* MA terms and *d* represents the differencing used to create a stationary time series

- Some special cases of ARIMA models

White noise	ARIMA(0,0,0)
Random walk	ARIMA(0,1,0) with no constant
Random walk with drift	ARIMA(0,1,0) with a constant
Autoregression	ARIMA(p,0,0)
Moving average	ARIMA(0,0,q)

ARIMA Forecasting

- Point forecasts can be calculated using the following three steps

- Expand the ARIMA equation so that y_t is on the left-hand side and all other terms are on the right

- Rewrite the equation by replacing t with T+h

- On the right-hand side of the equation, replace future observations with their forecasts, future errors with zero, and past errors with the corresponding residuals

ARIMA Model Identification based on ACF and PACF

- By considering the patterns of the autocorrelations and the partial autocorrelations, we can guess a reasonable model for the data

Model	Autocorrelations	Partial Autocorrelations
ARIMA(p,d,0)	Infinite. Tails off.	Finite. Cuts off after *p* lags.
ARIMA(0,d,q)	Finite. Cuts off after *q* lags.	Infinite. Tails off.
ARIMA(p,d,q)	Infinite. Tails off.	Infinite. Tails off.

The Box-Jenkins Method

- *Box - Jenkins Analysis* refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models

- The method is appropriate for time series of medium to long length (at least 50 observations)

Box-Jenkins Methodology

- Iteratively apply the following steps until Step 3 does not produce any improvement in the model:

- ***Step 1: Identification***
  - Using plots of the data, autocorrelations, partial autocorrelations, and other information, a class of simple ARIMA models is selected. This amounts to estimating appropriate values for p, d, and q

- ***Step 2: Estimation***
  - The phis and thetas of the selected model are estimated using maximum likelihood techniques, backcasting, etc., as outlined in Box-Jenkins (1976)

- ***Step 3: Diagnostic Checking***
  - The fitted model is checked for inadequacies by considering the autocorrelations of the residual series (the series of residual, or error, values).

Seasonal ARIMA

- This model is used when the time series exhibits some seasonality

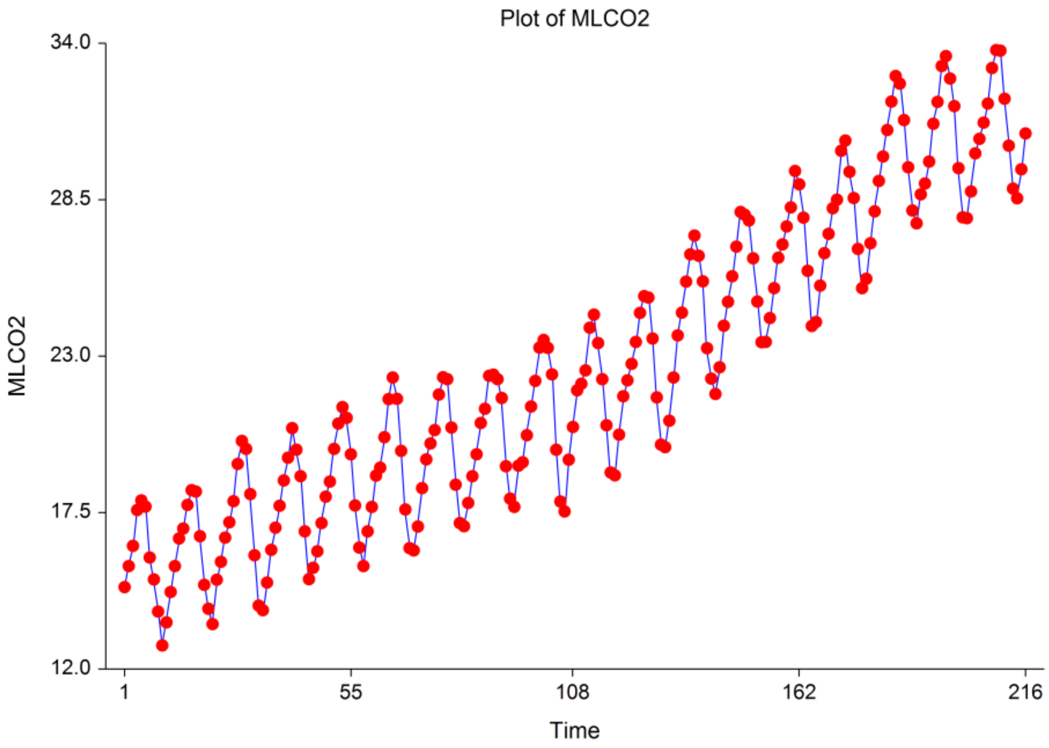
- Three additional parameters (P, D and Q) to account for the seasonal part of the model

ARIMA	$\underbrace{(p, d, q)}$	$\underbrace{(P, D, Q)_m}$
	$\uparrow$ Non-seasonal part of the model	$\uparrow$ Seasonal part of of the model

where m = number of observations per year.

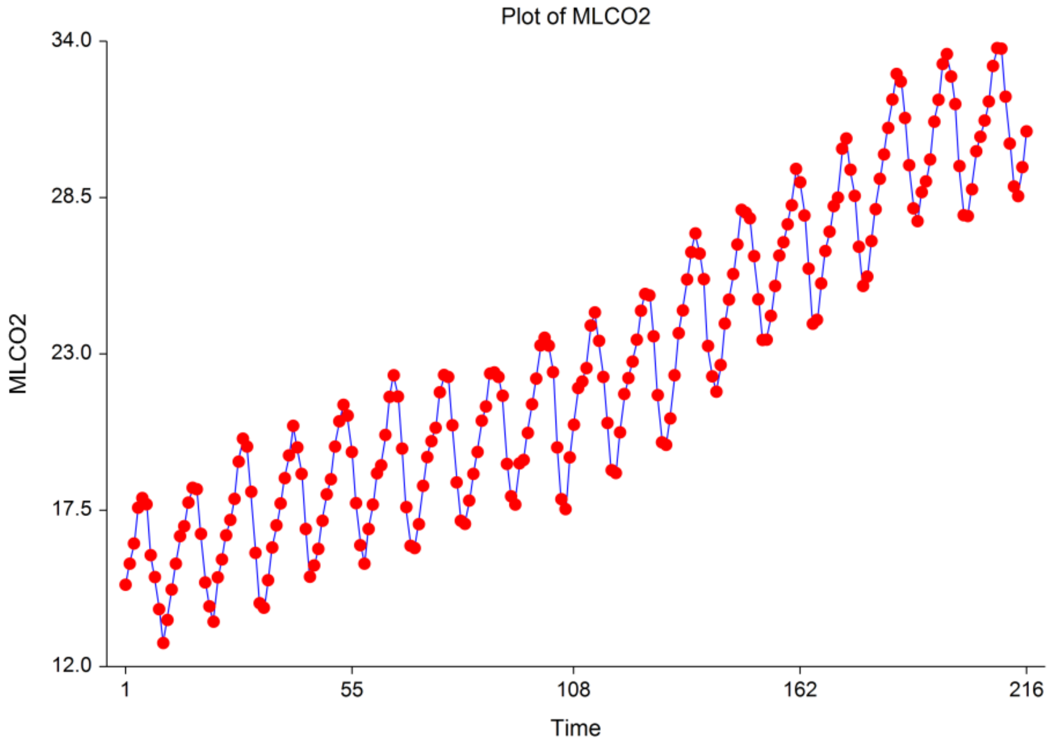
An Example of Box-Jenkins Methodology for ARIMA Modeling

- We will consider 216 monthly carbon dioxide measurements above Mauna Loa, Hawaii. The data was obtained from Newton (1988)

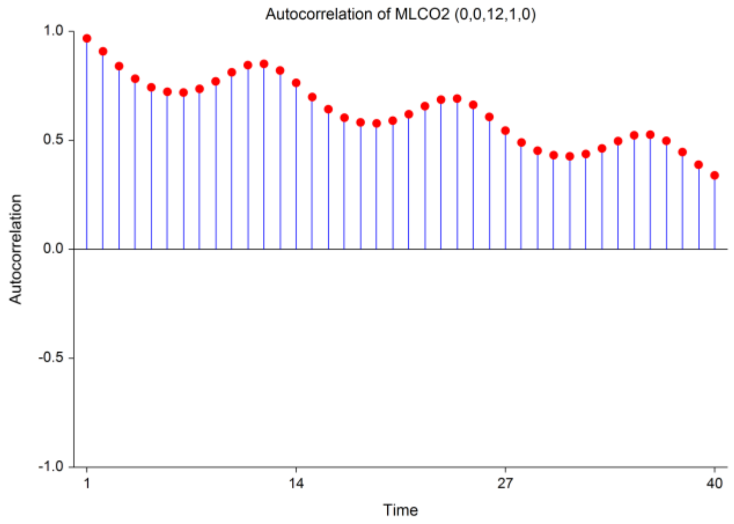


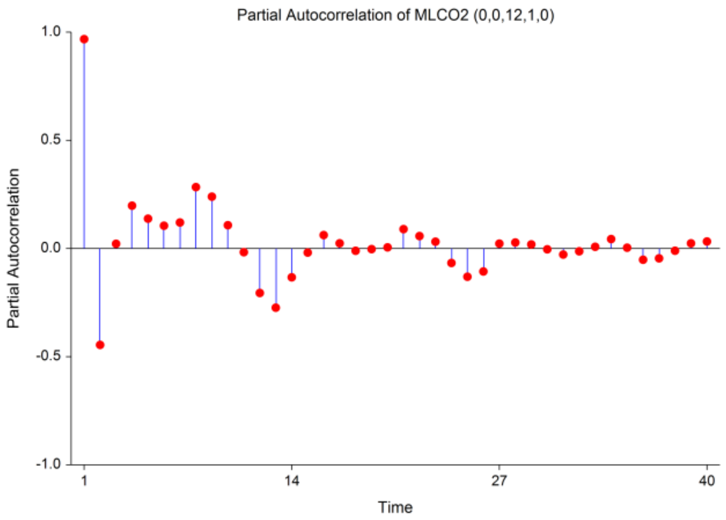
Quiz

➢ Is this data stationary? If not, what makes it non-stationary?



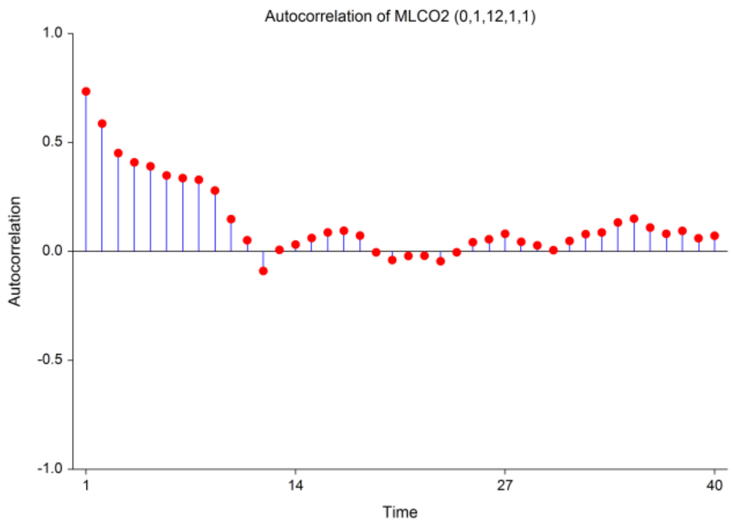
Step 1: Plot ACF and PACF

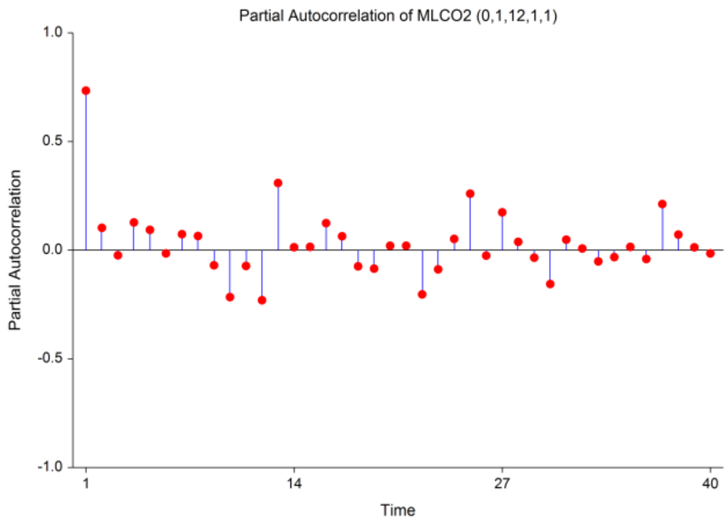




- Autocorrelations do not die out and they show a cyclical pattern. This points to non-stationarity in the data
- The partial autocorrelations point to a value of 2 for p
- **Caution:** first use differencing to make series stationary
- Because these are monthly data, we use seasonal difference of length twelve. We also remove the trend in the data

ACF and PACF for Seasonally Differenced Series





- The autocorrelations die out fairly quickly. The partial autocorrelations are large around lags one and twelve

- This suggests the multiplicative seasonal model: ARIMA(1,0,0) x $(1,1,0)_{12}$

Step 2: Estimation

Model Description Section

Series	MLCO2-TREND
Model	Regular(1,0,1) Seasonal(1,1,0) Seasons = 12
Trend Equation	(14.07418)+(7.830546E-02)x(date)
Observations	216
Iterations	13
Pseudo R-Squared	99.500042
Residual Sum of Squares	30.3262
Mean Square Error	0.1508766
Root Mean Square	0.3884284

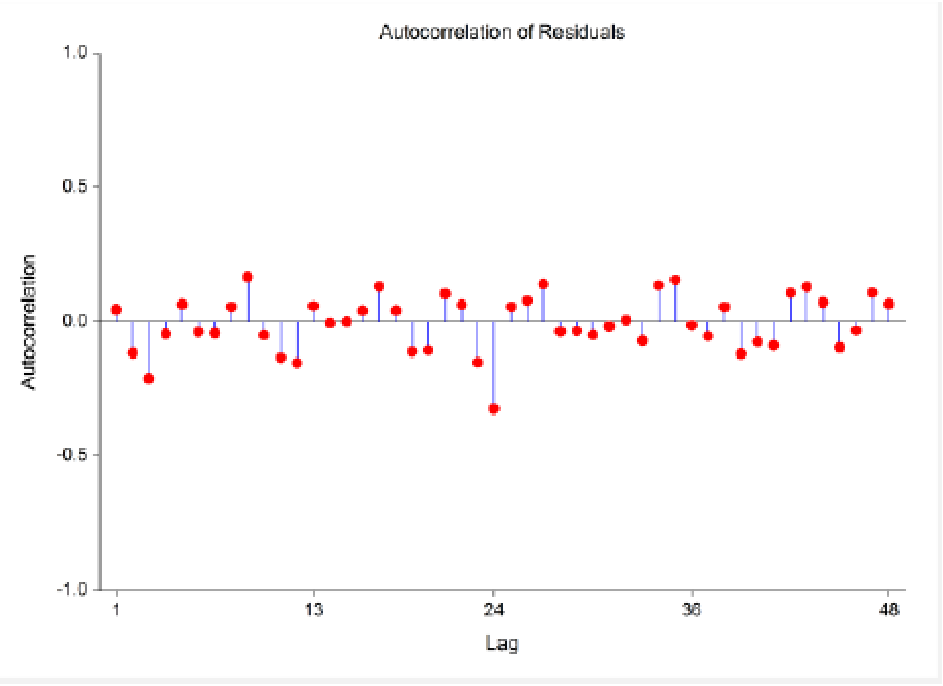
Model Estimation Section

Parameter Name	Parameter Estimate	Standard Error	T-Value	Prob Level
AR(1)	0.9836381	1.274416E-02	77.1834	0.000000
SAR(1)	-0.4927093	5.991305E-02	-8.2237	0.000000
MA(1)	0.3183001	6.915411E-02	4.6028	0.000004

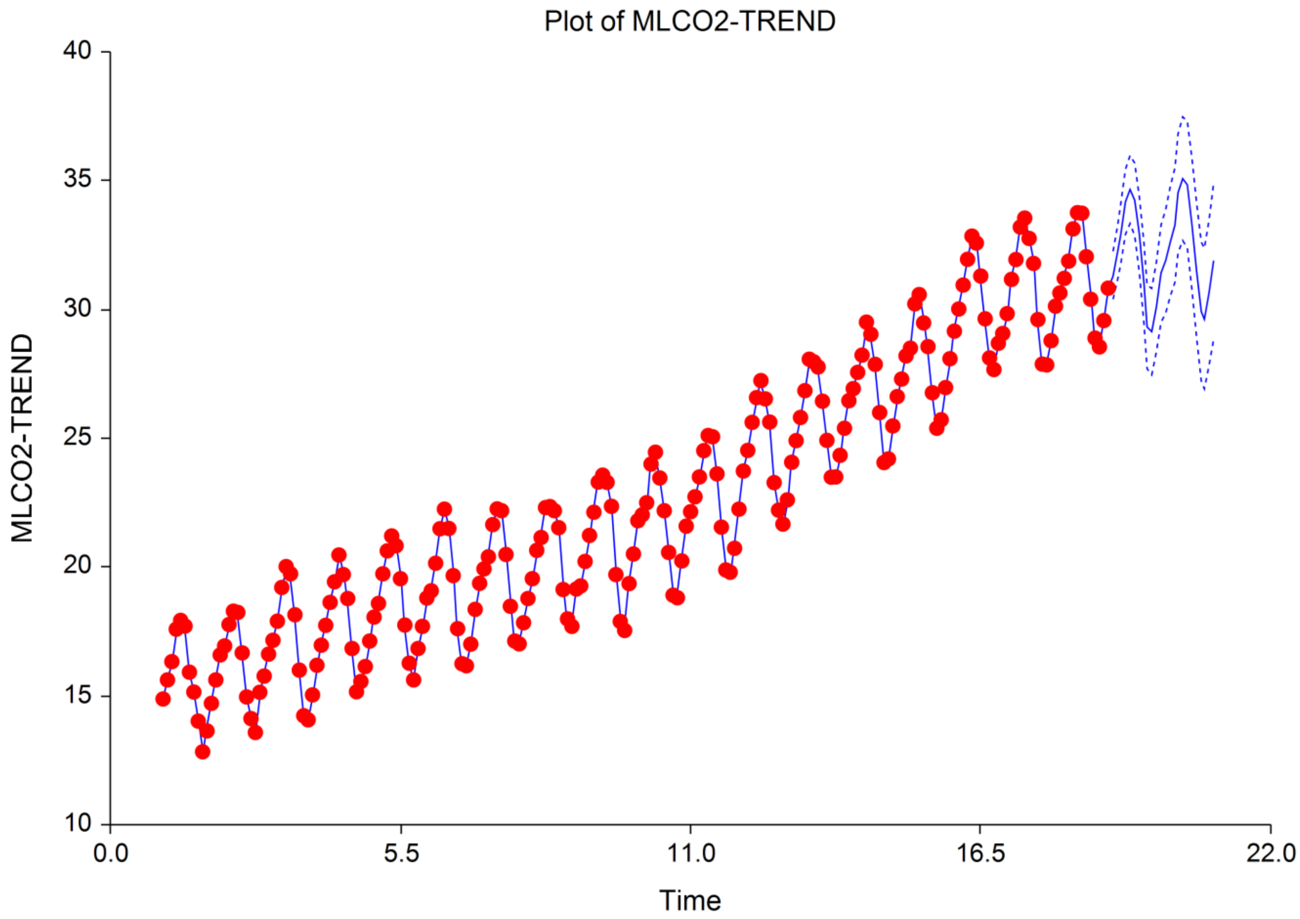
Step 3: Diagnostics Checking

- There appear to be some persistent autocorrelations at lag 25

- There may be room to further improve the model



Forecasts for ARIMA Model



ARIMA vs. ETS

- Linear exponential smoothing models are all special cases of ARIMA models
- Non-linear exponential smoothing models have no equivalent ARIMA counterparts
- All ETS models are non-stationary
- Some ARIMA models are stationary

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Model Evaluation

Evaluation

➢ Forecast error: Difference between forecast and actual values

$$ME = \frac{1}{N}\sum_{i=1}^{N}(A_i - F_i)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|A_i - F_i|$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(A_i - F_i)^2$$

$$Relative\ Error = 100 \times \frac{A_i - F_i}{A_i}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{A_i - F_i}{A_i}\right|$$

Table – Summary of Discussed Forecasting Metrics

	Absolute	**Relative**
Error Measures Bias	Mean Error (ME)	Mean Percentage Error (MPE)
Absolute Error Measures accuracy	Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)
Squared Error Measures accuracy, especially penalizing very bad forecast points	Mean Squared Error (MSE) Root-Mean Squared Error (RMSE)	Mean Squared Percentage Error (MSPE) Root-Mean squared Percentage Error (RMSPE)

Comparing Statistical and ML Models for Forecasting

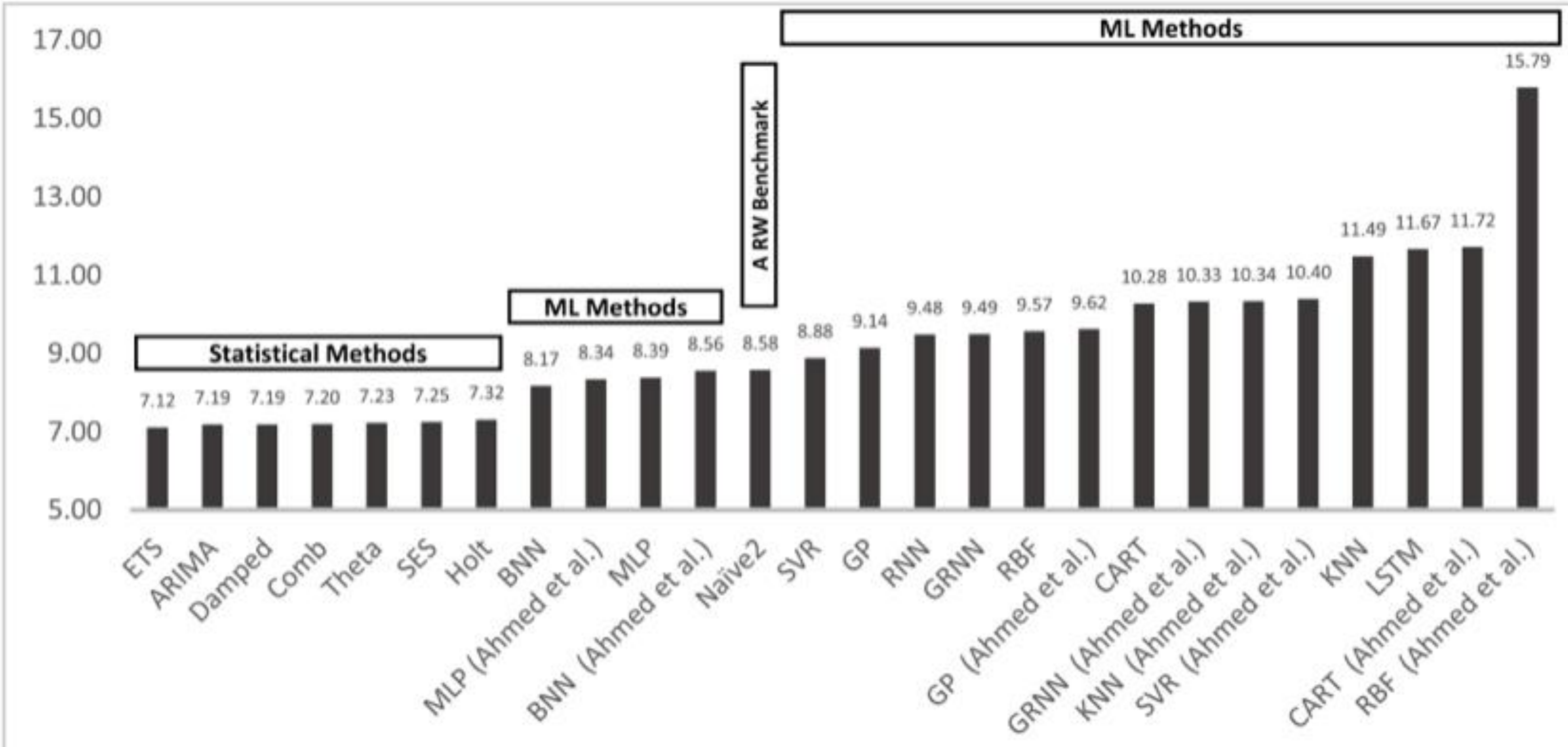


Fig 2. Forecasting performance (sMAPE) of the ML and statistical methods included in the study. The results are reported for one-step-ahead forecasts having applied the most appropriate preprocessing alternative.

*Statistical and Machine Learning forecasting methods: Concerns and ways forward

Resources

- Forecasting: Principles and Practice
  - Free version: https://otexts.com/fpp3/

- Time Series Analysis and Its Applications With R Examples

- Follow Rob Hyndman's blog https://robjhyndman.com/

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Hands-on Lab

ON-BRAND STATEMENT

FOR GENERAL USE

> What defines the students and faculty of the University of Washington? Above all, it's our belief in possibility and our unshakable optimism. It's a connection to others, both near and far. It's a hunger that pushes us to tackle challenges and pursue progress. It's the conviction that together we can create a world of good. And it's our determination to Be Boundless. Join the journey at **uw.edu**.