ISA 444: Business Forecasting

19: ARIMA Models

Fadel M. Megahed, PhD

Endres Associate Professor
Farmer School of Business
Miami University

@FadelMegahed
fmegahed
fmegahed@miamioh.edu

Automated Scheduler for Office Hours

Spring 2023

Quick Refresher from Last Class

ARMA Models: Models we considered may have three components, an autoregressive component (AR), and a moving average component (MA).

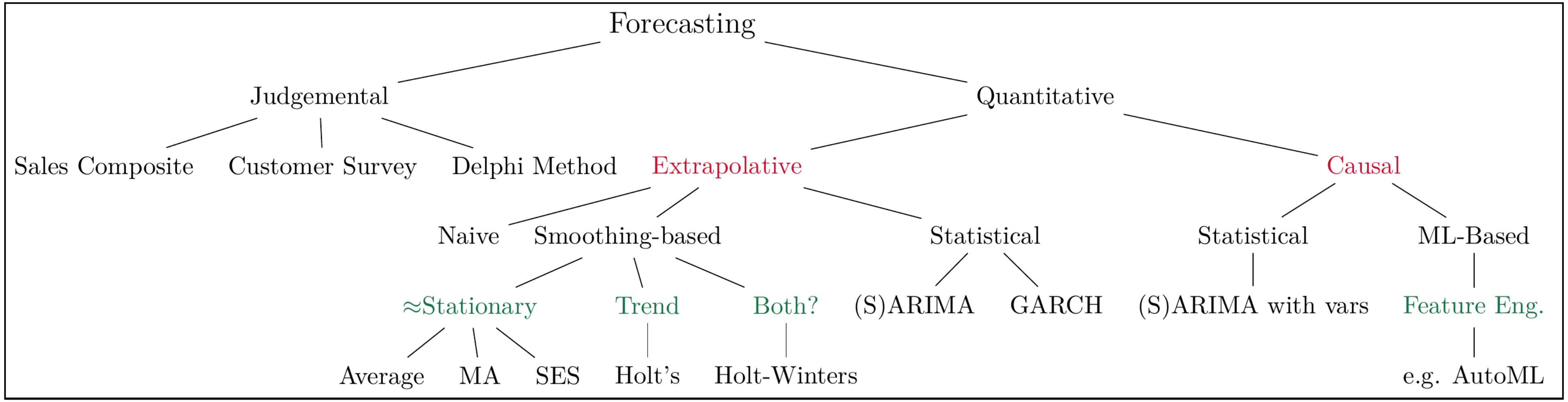
✅ Describe the behavior of the ACF and PACF of an AR(p) process.

✅ Describe the behavior of the ACF and PACF of an MA(q) process.

✅ Describe the behavior of the ACF and PACF of an ARMA (p,q) process.

✅ Fit an ARMA model to a time series, evaluate the residuals of a fitted ARMA model to assess goodness of fit, use the Ljung-Box test for correlation among the residuals of an ARMA model.

Overview of Univariate Forecasting Methods



A 10,000 foot view of forecasting techniques

Notes: My (incomplete) classification of **univariate** forecasting techniques, i.e., they exclude popular approaches used in multivariate time series forecasting.

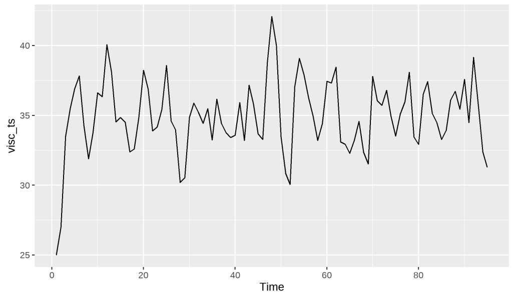
ARMA Processes

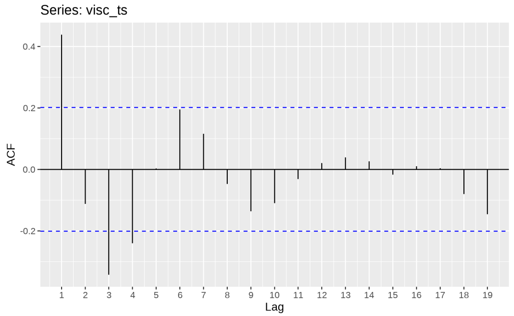
Last class, we discussed on how we capitalized on the ACF and PACF plots to identify which ARMA model can be used to model/approximate the observed time-series.

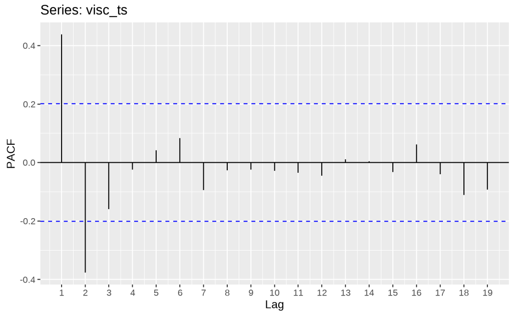
Model	ACF	PACF
AR(p)	Exponentially decays or damped sinusoidal pattern	Cuts off after lag p
MA(q)	Cuts off after lag q	Exponentially decays or damped sinusoidal pattern
ARMA(p,q)	Exponentially decays or damped sinusoidal pattern	Exponentially decays or damped sinusoidal pattern

Non-graded Activity: A 3rd Model to Visc?

```r
visc = readr::read_csv('../../data/viscosity.csv')

# Step 1: Plot the Data
visc_ts = ts(data = visc$y)
forecast::autoplot(visc_ts)

# Step 2: Stationary?
forecast::ndiffs(x = visc_ts)

# Step 3: Let us Look at the ACF and PACF plot
acf(x = visc_ts, plot = F) |>
  forecast::autoplot() +
  ggplot2::scale_x_continuous(
    breaks = scales::pretty_breaks(n = 25)
    )

pacf(x = visc_ts, plot = F) |>
  forecast::autoplot() +
  ggplot2::scale_x_continuous(
    breaks = scales::pretty_breaks(n = 25)
    )
```







In class, we examined AR(2) and MA(4) to model the viscosity data.
Use an AR(3) model to fit this dataset and comment on whether this model is suitable/reasonable to use.

Kahoot Competition #03

To assess your understanding and retention of the topics covered so far, you will **compete in a Kahoot competition (consisting of 7 questions)**:

- Go to https://kahoot.it/
- Enter the game pin, which will be shown during class
- Provide your first (preferred) and last name
- Answer each question within the allocated time window (**fast and correct answers provide more points**)

Winning the competition involves having as many correct answers as possible AND taking the shortest duration to answer these questions. The winner 🏆 of the competition will receive a $10 Starbucks gift card. Good luck!!!

P.S: The Kahoot competition will have **no impact on your grade**. It is a **fun** way of assessing your knowledge, motivating you to ask questions about topics covered that you do not have a full understanding of it, and providing me with some data to pace class.

Learning Objectives for Today's Class

- Explain how ARIMA models work when compared to ARMA models.
- Fit an ARIMA model to a time series, evaluate the residuals of a fitted ARMA model to assess goodness of fit, use the Ljung-Box test for correlation among the residuals of an ARIMA model.
- Describe AIC, AICc, and BIC and how they are used to measure model fit.
- Describe the algorithm used within the `auto.arima()` function to fit an ARIMA model.
- Describe the results of the `auto.arima()` function.

ARIMA Models

ARIMA Models for Nonstationary Processes

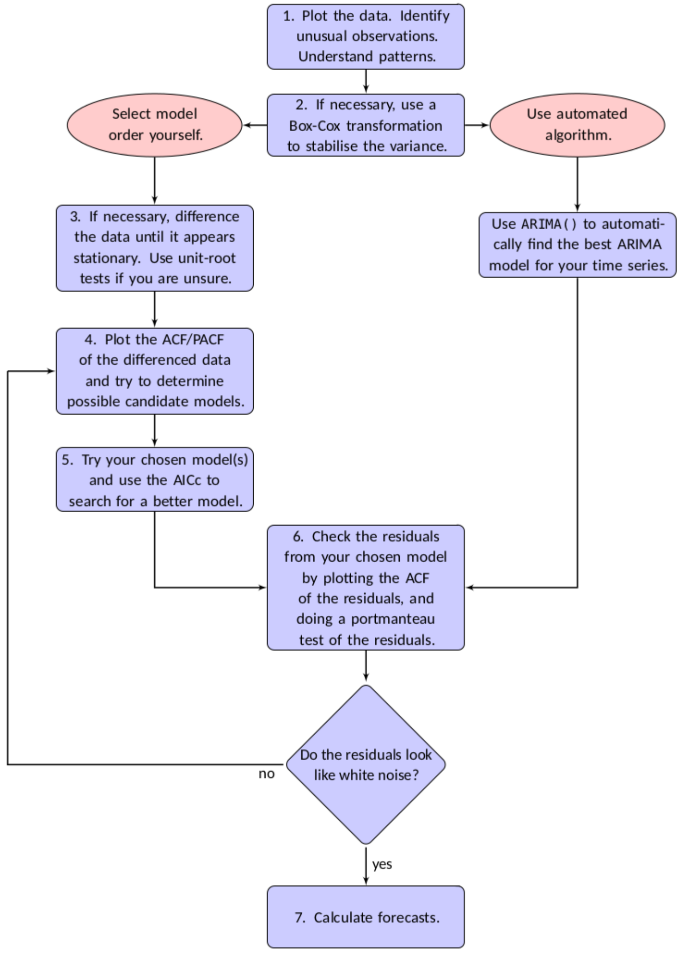
- When the time series is nonstationary, differencing can be used to transform the series.
    - The `forecast::ndiffs()` function can be used to determine: (a) whether the ts is stationary, and (b) the number of differences necessary to achieve stationarity.

Our 5-Step ARMA Fitting Process

- Plot the data over time.
- Do the data seem stationary? If necessary, conduct a test for stationarity.
- Once you can assume stationarity, find the ACF plot.
  - If the ACF plot cuts off, fit an MA(q), where $q =$ the cutoff point.
  - If the ACF plot dies down, find the PACF plot.
  - If the PACF plot cuts off, fit an AR(p) model, where $p =$ the cutoff point.
  - If the PACF plot dies down, fit an ARMA (p,q) model.
    - You must iterate through p and q using a guess and check method starting with ARMA(1,1) models -- increment each by 1.
- Evaluate the model residuals and consider the ACF and PACF of the residuals.
- If model fit is good, forecast future values.

Note: Often you will fit multiple models in Step 3 and compare models in Step 4 to select the best fit.

Extending Our 5-Step ARMA Fitting Process



Fitting an Appropriate ARIMA for Modeling GNP

In class, we will use the GNP Data to highlight how ARIMA models can be fit and used for forecasting.

Goodness of Fit Measures

Akaike Information Criterion (AIC)

- The AIC is a statistical measure used to **compare the goodness of fit of different statistical models to a given dataset**.
- AIC is based on the `parsimony` principle, favoring simpler models with fewer parameters over more complex ones (while still accounting for the model's ability to fit the data).
  - This helps to avoid overfitting, which occurs when a model is too complex and fits the noise in the data, rather than the underlying pattern.
- The **AIC formula** is given by: $AIC = 2k - 2ln(L)$, where
  - `k` is the number of parameters in the model (including the intercept, if applicable)
  - `L` is the likelihood of the model, measuring how well the model fits the observed data

In essence, the AIC balances the trade-off between the goodness of fit (measured by the log-likelihood) and the complexity of the model (measured by the number of parameters). **Lower AIC values indicate better model performance, so when comparing multiple models, the one with the lowest AIC is generally considered the best fitting, while also being the simplest**.

Corrected Akaike Information Criterion (AICc)

- The AICcis a modification of the AIC that provides a **more accurate model selection criterion when the sample size is small or the ratio of the sample size to the number of model parameters is low**.
- Like the AIC, the AICc is used to compare different statistical models and select the one that best fits the data while penalizing model complexity.
- $AICc = AIC + (2 * k * (k + 1))/(n - k - 1)$, where
    - `AIC` is the original Akaike Information Criterion: $AIC = 2k - 2ln(L)$
    - `k` is the number of parameters in the model (including the intercept, if applicable)
    - `n` is the sample size
- The **second term in the AICc formula represents a correction factor that accounts for the small sample size**. As the sample size (n) increases, the correction factor approaches zero, and the AICc converges to the original AIC. Therefore, in cases where the sample size is large, using the AICc instead of AIC won't make a significant difference.
- When comparing models, the one with the lowest AICc value is considered the best-fitting model while also being the simplest, similar to the AIC.

Bayesian Information Criterion (BIC)

- The BIC helps in model selection by balancing the trade-off between model complexity and goodness of fit. However, the **BIC penalizes model complexity more heavily than the AIC**.
- The formula for BIC is: $BIC = -2ln(L) + k * ln(n)$, where
  - L is the likelihood of the model, measuring how well the model fits the observed data
  - k is the number of parameters in the model (including the intercept, if applicable)
  - n is the sample size
- In the BIC formula, the first term represents the goodness of fit (measured by the log-likelihood), and the second term represents the penalty for model complexity.
- The penalty term is larger in the BIC than in the AIC, as it is proportional to the natural logarithm of the sample size instead of being a constant value.
- When comparing models, the one with the lowest BIC value is considered the best-fitting model while also being the simplest, similar to the AIC and AICc.
- The BIC might be preferred over the AIC or AICc, especially when there is **a preference for more parsimonious models or when dealing with large sample sizes**.

AIC, AICc and BIC Summary

Studies have shown that the:

- BIC does well at getting correct model in large samples.
- AICc tends to get correct models in smaller samples with a large number of parameters.

Why did we discuss these metrics today?

- They were printed with some of the models that we have examined in class.
- They are used with the `auto.arima()`, which comes from the forecast package.

The `auto.arima()` function

The `auto.arima()` function

The auto.arima() function can be used to automatically fit ARIMA models to a time series. It is a useful function, but it should be used with caution.

The `forecast::auto.arima()` function

- Uses "brute force" to fit many models and then selects the "best" based on a certain model criterion
- Works best when the data are stationary, but can be used with nonstationary data
- Tends to overfit the data
- Should always be used as a starting point for selecting a model and all models derived from the `auto.arima()` function should be properly vetted and evaluated.

The `forecast::auto.arima()` function combines:

- Unit root tests (KPSS by default)
- Minimization of AICc to obtain an `ARIMA(p, d, q)` model using the following algorithm:

The `auto.arima()` function

- Determine the number of differences, d, using a sequence of KPSS tests.
  - Determine p and q by minimizing AICc after differencing the data d times. Rather than considering all possible p and q combinations, a stepwise approach is taken.
    - The best initial model with lowest AICc is selected from the following four:
      - ARIMA(2,d,2),
      - ARIMA(0,d,0),
      - ARIMA(1,d,0), and
      - ARIMA(0,d,1).
      - *If d=0, then a constant, c, is included. If $d \geq 1$, then the constant is set to 0. The results of this step is called the current model.*
    - Variations on the current model are considered by
      - Vary p and/or q from current model by ± 1
      - Include/exclude c from current model.
      - The best model considered so far (current or one of variations) becomes the *new current model.*
    - Repeat previous step until no lower AICc can be found.

Live Coding

Let us examine how the `forecast::auto.arima()` can be used to model and forecast the GNP Data.

Recap

Summary of Main Points

By now, you should be able to do the following:

- Explain how ARIMA models work when compared to ARMA models.
- Fit an ARIMA model to a time series, evaluate the residuals of a fitted ARMA model to assess goodness of fit, use the Ljung-Box test for correlation among the residuals of an ARIMA model.
- Describe AIC, AICc, and BIC and how they are used to measure model fit.
- Describe the algorithm used within the `auto.arima()` function to fit an ARIMA model.
- Describe the results of the `auto.arima()` function.

Things to Do to Prepare for Next Class

- Go through the slides, examples and make sure you have a good understanding of what we have covered.
- Read Chapters $9.1, 9.3 - 9.7$ in Forecasting: Principles and Practice.
- Complete Assignment 13.