GENERAL ASSEMBLY

TIME SERIES ANALYSIS: ARMA

Joseph Nelson + Matt Brems

Data Science Immersive, GA DC

LEARNING OBJECTIVES

- Describe the difference between weakly and strongly stationary stochastic processes.
- Understand, describe, and implement autoregressive models.
- Understand, describe, and implement moving average models.
- Understand, describe, and implement ARMA models.

OPENING: TIME SERIES DATA

- A univariate time series is a sequence of measurements of the same variable collected over time.
  - Formally, we call this a stochastic process and might denote these measurements as random variables. $\{Y_t : t = 0, 1, \ldots, n\}$.
  - Most often (but not always) these measurements will be made at regular time intervals.

STATIONARY PROCESSES

STATIONARY PROCESSES

- A stochastic process $\{Y_t: t = 0, 1, \ldots\}$ is strongly stationary if
  - $f\left(Y_{t_0}, Y_{t_1}, \ldots, Y_{t_n}\right) = f\left(Y_{t_0+k}, Y_{t_1+k}, \ldots, Y_{t_n+k}\right)$ for all t_i and k.
  - More practically:
    - $\mu_{t_i} = \mu_{t_i+k}$
    - $\mathrm{Cov}(Y_t, Y_{t+k}) = \mathrm{Var}(Y_k)$
    - $\mathrm{Corr}(Y_t, Y_{t+k}) = \frac{\mathrm{Var}(Y_k)}{\mathrm{Var}(Y_0)}$

STATIONARY PROCESSES

- A stochastic process $\{Y_t: t = 0, 1, \dots\}$ is strongly stationary if
  - $f(Y_{t_0}, Y_{t_1}, \dots, Y_{t_n}) = f(Y_{t_0+k}, Y_{t_1+k}, \dots, Y_{t_n+k})$ for all t_i and k.
  - More practically:
    - $\mu_{t_i} = \mu_{t_i+k}$
    - $\mathrm{Cov}(Y_t, Y_{t+k}) = \mathrm{Var}(Y_k)$
    - $\mathrm{Corr}(Y_t, Y_{t+k}) = \frac{\mathrm{Var}(Y_k)}{\mathrm{Var}(Y_0)}$

- A stochastic process $\{Y_t: t = 0, 1, \dots\}$ is weakly stationary if
  - $\mu_{t_i} = c$ for all t_i
  - $\mathrm{Cov}(Y_t, Y_{t+k}) = \mathrm{Cov}(Y_0, Y_k)$

STATIONARY PROCESSES

- Stationarity means that some characteristic of the distribution of a stochastic process does not depend on the spatial location, only the displacement (distance) between the times/locations.
- Time series data is often assumed to be stationary. (While not always valid, you can include other components so that you are left with a stationary component.)

- Example of stationary processes:
    - Rainfall: The relationship between rainfall in 2015 and 2017 is the same as the relationship between rainfall in 2014 and 2016, which is the same as the relationship between rainfall in 2013 and 2015, and so on.
        - Stationarity means that the relationship between the rainfall in one year and rainfall in another year depends only on the “distance” or “amount of time” between the two and nothing else.

STATIONARY PROCESSES

- Stationarity is often a convenient condition to have because it has lots of nice properties that allow results to “collapse down.”
    - Weak stationarity is required for moving average models! (We’ll get into that later.)

AUTOREGRESSIVE MODELS (AR)

AUTOREGRESSIVE MODELS

- Autoregressive (AR) models use data from previous time points to predict the next time point.

AUTOREGRESSIVE MODELS

- Autoregressive (AR) models use data from previous time points to predict the next time point.

- These are essentially regression models where the predictors are previous timepoints of the outcome.

AUTOREGRESSIVE MODELS

- Autoregressive (AR) models use data from previous time points to predict the next time point.

- These are essentially regression models where the predictors are previous timepoints of the outcome.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon$$

AUTOREGRESSIVE MODELS

- Autoregressive (AR) models use data from previous time points to predict the next time point.

- These are essentially regression models where the predictors are previous timepoints of the outcome.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon$$

- We say this model is $\mathrm{AR}(p)$ because it includes the previous p values of Y.
    - The most common autoregressive model is $\mathrm{AR}(1)$.

AUTOREGRESSIVE MODELS

- What might be our biggest concern with fitting an $\mathrm{AR}(p)$ model?

AUTOREGRESSIVE MODELS

- What might be our biggest concern with fitting an $\mathrm{AR}(1)$ model?
- A time series with high autocorrelation implies that the data is highly dependent on previous values – therefore, an autoregressive model would perform well!

AUTOREGRESSIVE MODELS

- What might be our biggest concern with fitting an $\mathrm{AR}(1)$ model?
- A time series with high autocorrelation implies that the data is highly dependent on previous values – therefore, an autoregressive model would perform well!
- Autoregressive models are useful for learning "falls" or "rises" in our series.
    - Typically this model type is useful for small-scale trends.

AUTOREGRESSIVE MODELS

- Recall an AR(p) model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon$$

- Note that we could instead write this as:

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \varepsilon$$

MOVING AVERAGE MODELS

- Moving average models take previous error terms as inputs.
  - By contrast: Autoregressive (AR) models use data from previous time points to predict the next time point.

MOVING AVERAGE MODELS

- Moving average models take previous error terms as inputs.
  - By contrast: Autoregressive (AR) models use data from previous time points to predict the next time point.

- Moving average models predict the next value based on the overall average and how incorrect our previous predictions were.
  - This is useful for modeling a sudden occurrence – like something going out of stock and thus sales are affected, or Donald Trump tweets about an organization and its stock price falls.

MOVING AVERAGE MODELS

- As with AR models, we have an order term q and we call our moving average model $\mathrm{MA}(q)$.

$$Y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q}$$

- In this case, μ will be the expected value of Y_t.
  - This can be the global mean or, if you are assuming the values of Y follow a distribution, the theorized mean for that distribution.

MOVING AVERAGE MODELS

- As with AR models, we have an order term q and we call our moving average model $\text{MA}(q)$.

$$Y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q}$$

- MA models require a more complex fitting procedure than AR models:
  - Iteratively fit model.
  - Compute errors.
  - Refit.

MOVING AVERAGE MODELS

- Note that MA models include the mean of the time series. Thus the behavior of an MA model is characterized by random jumps around the mean value.

$$Y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q}$$

MOVING AVERAGE MODELS

- Note that MA models include the mean of the time series. Thus the behavior of an MA model is characterized by random jumps around the mean value.

$$Y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \cdots + \beta_q \varepsilon_{t-q}$$

- Note that we could instead write this as:

$$Y_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}$$

MOVING AVERAGE MODELS ≠ MOVING AVERAGE

- NOTE: Moving Average ≠ Moving Average <u>Models</u>!!!
- Moving average <u>models</u> take, as inputs, previous errors.
- Moving averages are not models and just average recent Y values!

ARMA MODELS

ARMA MODELS

- ARMA(p, q) models combine the autoregressive and moving average models.

$$Y_t = c + \varepsilon_t + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-i}$$

$$\hat{Y}_t = c + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-i}$$

- Y_t is the “true value” and includes ε_t whereas $\hat{Y}_t$ is the “fitted value.”
- The constant c combines β_0 from the AR model and μ from the MA model.

ARMA MODELS

$$Y_t = c + \varepsilon_t + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-i}$$

$$\hat{Y}_t = c + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-i}$$

- This slowly incorporates changes in preferences, tastes, and patterns. We account for sudden changes based on random events as well as how we expect things to change over time given past events.