

# ISA 444: Business Forecasting

## 05 - Basic Tools for Time Series Analysis

Fadel M. Megahed

Associate Professor  
Department of Information Systems and Analytics  
Farmer School of Business  
Miami University  
Email: [fmegahed@miamioh.edu](mailto:fmegahed@miamioh.edu)  
Office Hours: [Click here to schedule an appointment](#)

Spring 2021

# Outline

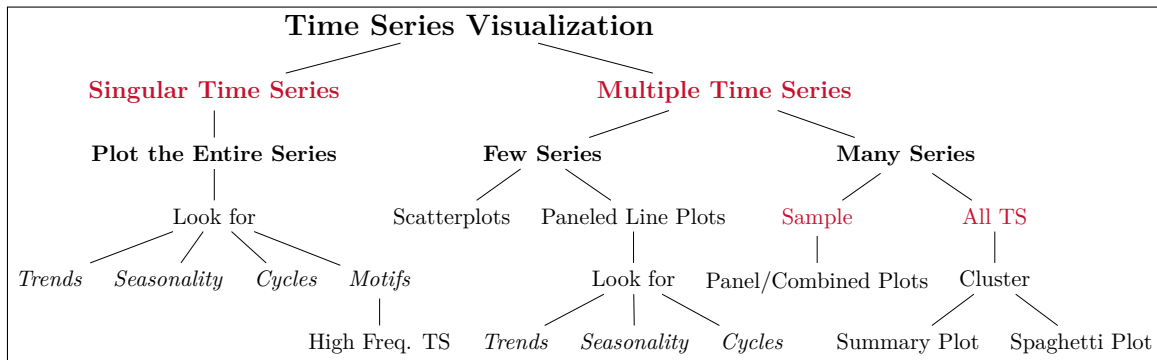
- 1 **Preface**
- 2 Transformations
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals
- 5 Recap

# What we Covered Last Week

## Main Learning Outcomes

- ✓ Explain different goals for visualizing time series data
- ✓ Identify an appropriate chart for a specific time series data visualization goal
- ✓ Use software to construct charts of interest
- ✓ Use numerical summaries to describe a time series.
- ✗ Apply transformations to a time series.

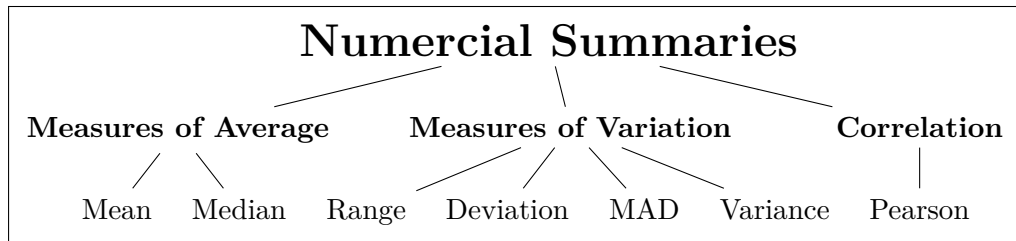
# Recap: A Structured Approach for Time Series Visualization



A Potential Framework for Time Series Visualization.<sup>1</sup>

<sup>1</sup>This is my best attempt to improve on the general advice provided in the previous slide. Many of the suggestions, presented in this flow chart, stem from my past and current research/consulting collaborations. They are by no means a comprehensive list of everything that you can do.

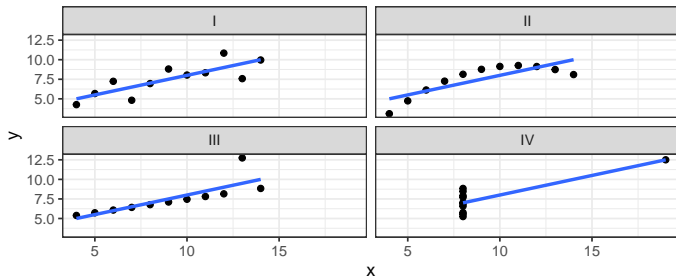
# Recap: Numerical Summaries



An overview of the numerical summaries discussed last class.

# Recap: Numerical Summaries do NOT Replace Visuals

set	x.mean	x.sd	y.mean	y.sd	corr
I	9.00	3.32	7.50	2.03	0.82
II	9.00	3.32	7.50	2.03	0.82
III	9.00	3.32	7.50	2.03	0.82
IV	9.00	3.32	7.50	2.03	0.82



# Learning Objectives for Today's Class

## Main Learning Outcomes

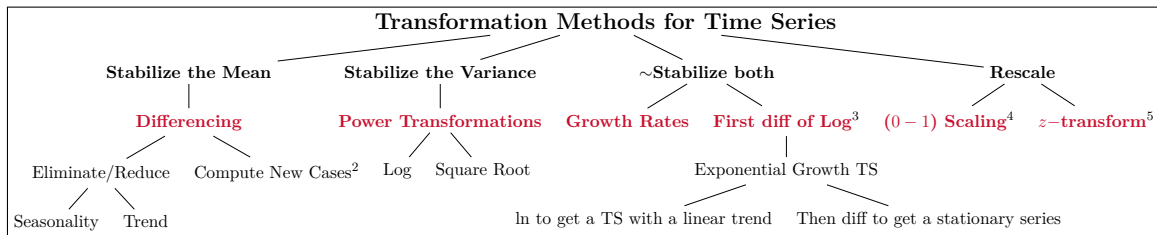
- Apply transformations to a time series.
- Apply and interpret measures of forecast accuracy.
- Interpret prediction intervals for a simple forecast.

# Outline

- 1 Preface
- 2 Transformations**
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals
- 5 Recap



# Guidelines for Transforming Time-Series Data



A classification of common transformation approaches for time series data.<sup>6</sup>

<sup>2</sup>The [COVID19 package](#) returns cumulative cases, i.e. a first difference  $\rightarrow$  new confirmed cases.

<sup>3</sup>First difference of LOG  $\cong$  percentage change. This is almost exact if the percentage change is small, but for larger percentage changes, it may differ greatly (see [here for more details](#)).

<sup>4</sup>Rescaling of the data from the original range so that all values are within the range of 0 and 1. Mathematically, speaking this can be achieved by calculating  $y_t = \frac{x_t - \min}{\max - \min}$ .

<sup>5</sup>One can normalize a time-series by  $z_t = \frac{x_t - \mu}{\sigma}$ .

<sup>6</sup>My (incomplete) attempt to provide you with a taxonomy for time series data transformations.

# Differencing

The change in the time series from one period to the next is known as the (first) difference. It can be computed as follows:

$$DY_t = Y_t - Y_{t-1}$$

$$\xrightarrow{\text{in R}} \text{variableName} - \text{lag}(\text{variableName}, x = 1)$$
(1)

Differences can be computed by capitalizing on `mutate()` and `lag()`.

symbol	date	adjusted	DYt
ADA-USD	2021-02-02	0.43	
ADA-USD	2021-02-03	0.44	0.02
ADA-USD	2021-02-04	0.44	-0.00
ADA-USD	2021-02-05	0.54	0.10
ADA-USD	2021-02-06	0.63	0.09

# Differencing for Seasonal Data

If your data exhibits a seasonal pattern, as illustrated in Slides 15-17 in [03-Time-Series-Plots.pdf](#), you should employ a **seasonal differencing approach**, you should subtract the difference between an observation and the previous observation from the same season. Let  $m$  denote the number of seasons, e.g.  $m = 4$  for quarterly data. In such a case, the seasonal difference is computed as follows:

$$DY_{t-m} = Y_t - Y_{t-m}$$

$$\xrightarrow{\text{in R}} \text{variableName} - \text{lag}(\text{variableName}, x = m) \quad (2)$$

## Growth Rates: The Formulation

In the absence of seasonality, the growth rate for a time series is given by

$$GY_t = 100 \frac{Y_t - Y_{t-1}}{Y_{t-1}} \quad (3)$$

In the presence of seasonality (with period =  $m$ ), the growth rate for a time series is given by

$$GY_t = 100 \frac{Y_t - Y_{t-m}}{Y_{t-m}} \quad (4)$$

## Growth Rates in Practice – a Non-Graded Class Activity

- **Question 1:** Let us say that an investor purchased 10 stocks of \$GME, on 2021-01-29, at \$325/stock. The next trading day, 2021-02-01, the \$GME stock closed at \$225. Compute the growth rate in their portfolio worth (assuming it only has the GME stock) over this time period.
- Let us say that the growth rate,  $GY_t = -g$ . Now let us assume that the \$GME stock went up by  $g$  (i.e., if it went down 10%, it increased by 10% over the next trading day). What is the value of the investor's portfolio by stock market closing on 2021-02-02?
  - Provide the answer to both computational questions on [Canvas](#).

## The Log Transform [1]

The log transformation can be computed as follows:

$$L_t = \ln(Y_t) \quad (5)$$

Note that the `log()` in R takes the natural logarithm as its default base, i.e., would transform a variable/statistic based on the above equation.

The reverse transformation using the exponential function is:

$$e^{L_t} = e^{\ln(Y_t)} = Y_t \quad (6)$$

The first difference in logarithms represents the logarithm of the ratio:

$$L_t = \ln\left(\frac{Y_t}{Y_{t-1}}\right) = \ln(Y_t) - \ln(Y_{t-1}) \quad (7)$$

# The Log Transform [2]

- The primary purpose of the log transform is to **convert exponential growth into linear growth**.
- The transform often has the **secondary purpose of balancing the variance**.
- Difference in logs and growth rate transformations produce similar results and interpretations.

# A Walk through Statistical Transformations Using R [1]

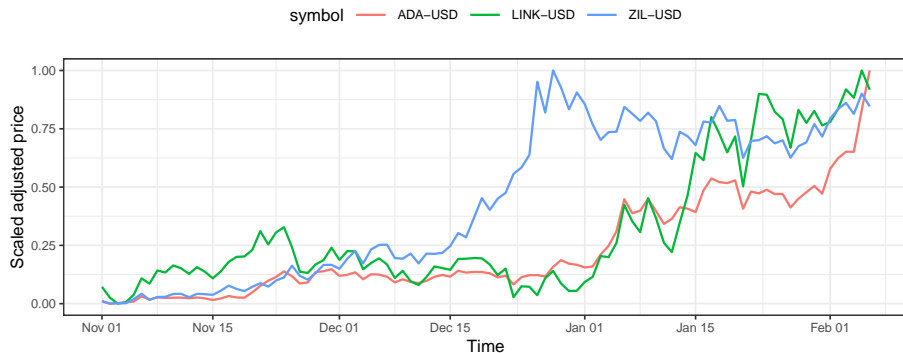
In this live coding session, we will capitalize on the `mutate()` from [tidyverse](#) to create transformations for multiple time series. Specifically, we will use the `tq_get()` from [tidyquant](#) to extract data about the following cryptocurrencies (a) [Cardano](#) (\$ADA), (b) [Chainlink](#) (\$LINK), and (c) [Zilliqa](#) (\$ZIL). We will compute:

- Growth Rates
- Natural log
- Log Differences
- $[0 - 1]$  Scaling

Obviously, we will have to ensure that these transformations are computed for each coin separately. For the purpose of this activity, let us extract the data from 2020-11-01 to 2021-02-06.

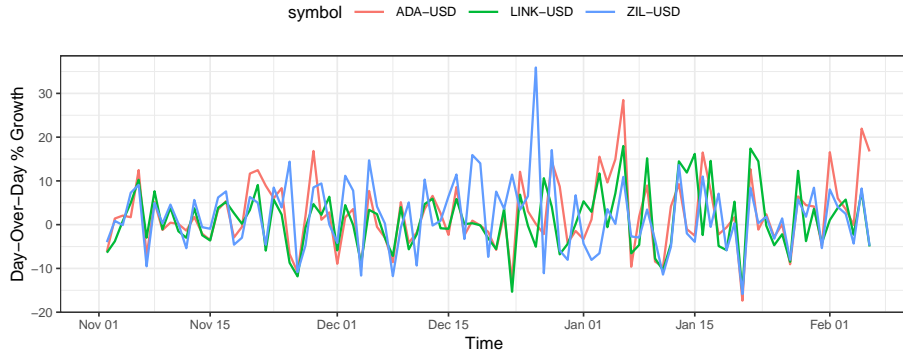


# A Walk through Statistical Transformations Using R [2]



**Question:** What insights can you get from this chart?

# A Walk through Statistical Transformations Using R [3]

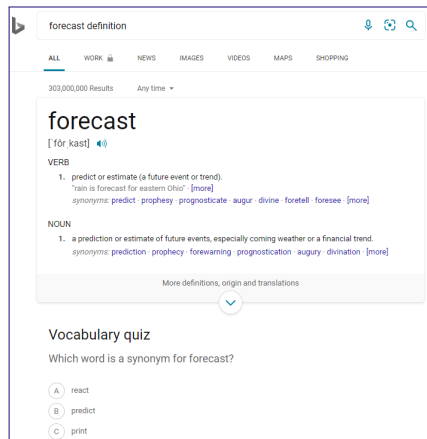


**Question:** What insights can you get from this chart?

# Outline

- 1 Preface
- 2 Transformations
- 3 Measures of Forecast Accuracy**
- 4 Prediction Intervals
- 5 Recap

# Recap: Definition of Forecast



The screenshot shows a Bing search result for the term "forecast". The search bar at the top contains "forecast definition". Below the search bar, there are tabs for "ALL", "WORK", "NEWS", "IMAGES", "VIDEOS", "MAPS", and "SHOPPING". The "ALL" tab is selected. The search results show "303,000,000 Results" and "Any time" filter. The main definition for "forecast" is displayed, including its pronunciation [fôr kast], part of speech (VERB), and a list of definitions. The first definition for the verb is "predict or estimate (a future event or trend)", with an example "rain is forecast for eastern Ohio" and synonyms: predict, prophesy, prognosticate, augur, divine, foretell, foresee. The second definition for the noun is "a prediction or estimate of future events, especially coming weather or a financial trend", with synonyms: prediction, prophecy, forewarning, prognostication, augury, divination. Below the definitions is a button labeled "More definitions, origin and translations". At the bottom, there is a "Vocabulary quiz" section with the question "Which word is a synonym for forecast?" and three options: A react, B predict, and C print.

forecast definition

ALL WORK NEWS IMAGES VIDEOS MAPS SHOPPING

303,000,000 Results Any time

**forecast**  
[fôr kast]

VERB

1. predict or estimate (a future event or trend).  
"rain is forecast for eastern Ohio" [more]  
synonyms: predict · prophesy · prognosticate · augur · divine · foretell · foresee [more]

NOUN

1. a prediction or estimate of future events, especially coming weather or a financial trend.  
synonyms: prediction · prophecy · forewarning · prognostication · augury · divination [more]

More definitions, origin and translations

Vocabulary quiz

Which word is a synonym for forecast?

A react  
B predict  
C print

The definition of the term "forecast" as obtained from Bing/Merriam-Webster.

# A Naïve Forecast

- A naïve forecast for an observation,  $Y_t$ , is the observation prior,  $Y_{t-1}$ .
- For some types of time series (e.g. Random Walks), a naïve forecast is the best possible forecast one can make<sup>7</sup>.
- In the case of seasonal data, a naïve forecast could be the observation from the prior period.
  - For example, in the case of monthly data, the naïve forecast for the observation  $Y_{Jan2018}$  could be  $Y_{Jan2017}$ . In this case, we would denote the frequency,  $m=12$ , and the naïve forecast for  $Y_t$  is the observation  $m$  periods prior, or  $Y_{t-m}$ .

---

<sup>7</sup>Slide is from [Dr. Allison Jones-Farmer's](#) lecture notes, Miami University, Spring 2020.

# Measures of Forecast Accuracy

The measures of accuracy we will discuss all deal with the difference between the actual observed value ( $Y_t$ ) and the forecasted value ( $F_t$ ) at time  $t$ . In order to measure forecast accuracy, we assume we have  $m$  actual values available, thus we have  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+m}$  and forecasts  $F_{t+1}, F_{t+2}, \dots, F_{t+m}$ . This is important because we will be averaging the forecast errors over  $m$ .<sup>8</sup>

## Forecast Error:

$$e_{t+i} = Y_{t+i} - F_{t+i}. \quad (8)$$

---

<sup>8</sup>Note that your textbook discusses rolling forecast origins. This is important, but we will save this discussion for later in the semester. For now, assume the forecast origin is fixed (i.e., we are only interested in the one-period ahead forecast).

# Measures Reflecting “Average” Forecast Performance

**Mean Error:**

$$ME = \frac{\sum_{i=1}^m e_{t+i}}{m}. \quad (9)$$

**Mean Percentage Error:**

$$MPE = \frac{100}{m} \sum_{i=1}^m \frac{e_{t+i}}{Y_{t+i}}. \quad (10)$$

# Computing Measures of “Average” Forecast Performance

symbol	date	adjusted	naiveFC	e	PE
ADA-USD	2021-01-31	0.34			
ADA-USD	2021-02-01	0.40	0.34	0.06	14.21
ADA-USD	2021-02-02	0.43	0.40	0.02	5.68
ADA-USD	2021-02-03	0.44	0.43	0.02	3.43
ADA-USD	2021-02-04	0.44	0.44	-0.00	-0.08
ADA-USD	2021-02-05	0.54	0.44	0.10	18.00
ADA-USD	2021-02-06	0.63	0.54	0.09	14.34

The ME and MPE are equal to 0.05 and 9.26% , respectively.

## Comments:

- We are forecasting the adjusted/closing price of the \$ADA stock. Given that this crypto, trading is 24/7 (i.e., the **next trading day always equals the next day**).
- The naïve forecast is the `lag(1)` of the series; thus, the forecast error is the `diff(1)`.
- A **seasonal naïve forecast** can be computed using `lag(m)` or the `snaive()` from the `fpp2` package.



# Measures Reflecting “Variability” in Forecast Performance

**Absolute Forecast Error:**

$$|e_{t+i}| = |Y_{t+i} - F_{t+i}|. \quad (11)$$

**Squared Forecast Error:**

$$(e_{t+i})^2 = (Y_{t+i} - F_{t+i})^2. \quad (12)$$

**Mean Absolute Error:**

$$MAE = \frac{\sum_{i=1}^m |e_{t+i}|}{m}. \quad (13)$$

**Root Mean Squared Error:**

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (e_{t+i})^2}{m}}. \quad (14)$$

# Measures Reflecting “Relative” Forecast Performance

**Mean Absolute Percentage Error:**

$$MAPE = \frac{100}{m} \sum_{i=1}^m \frac{|e_{t+i}|}{m}. \quad (15)$$

**Relative Mean Absolute Error:**

$$RelMAE = \frac{\sum_{i=1}^m |e_{t+i}|}{\sum_{i=1}^m |Y_{t+i} - Y_{t+i-1}|}. \quad (16)$$

**Thiel's U:**

$$U = \sqrt{\frac{\sum_{i=1}^m (e_{t+i})^2}{\sum_{i=1}^m (Y_{t+i} - Y_{t+i-1})^2}}. \quad (17)$$

# An Overview of Computing these Measures in R [1]

```
pacman::p_load(tidyquant, magrittr, fpp2, xtable)

cardano = tq_get("ADA-USD", from = "2021-01-31", to = "2021-02-06") %>%
  select(symbol, date, adjusted)
cardano %<>% mutate(naiveFC = lag(adjusted))

e = cardano$adjusted - cardano$naiveFC
ME = mean(e, na.rm=T)
RMSE = mean(e^2, na.rm=T) %>% sqrt()
MAE = abs(e) %>% mean(na.rm=T)
MPE = 100 * mean(e/cardano$adjusted, na.rm=T)
MAPE= 100 * mean(abs(e)/cardano$adjusted, na.rm=T)
```

# An Overview of Computing these Measures in R [2]

```
E = c(ME, RMSE, MAE, MPE, MAPE)
names(E) = c("ME", "RMSE", "MAE", "MPE", "MAPE")
round(E, 2) %>% print()
```

```
##      ME RMSE  MAE  MPE MAPE
## 0.05 0.06 0.05 9.26 9.29
```

```
# Alternatively, we could have just computed it using the fpp2 package
accuracy(object = cardano$naiveFC, # forecast object is the first argument
          cardano$adjusted) %>% round(2)
```

```
##              ME RMSE  MAE  MPE MAPE
## Test set 0.05 0.06 0.05 9.26 9.29
```

# Outline

- 1 Preface
- 2 Transformations
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals**
- 5 Recap

# Point vs Interval Forecasts

- **Point Forecasts:** future observations for which we report a single forecast observation.
- **Interval Forecast:** a range of values that are reported to forecast an outcome.

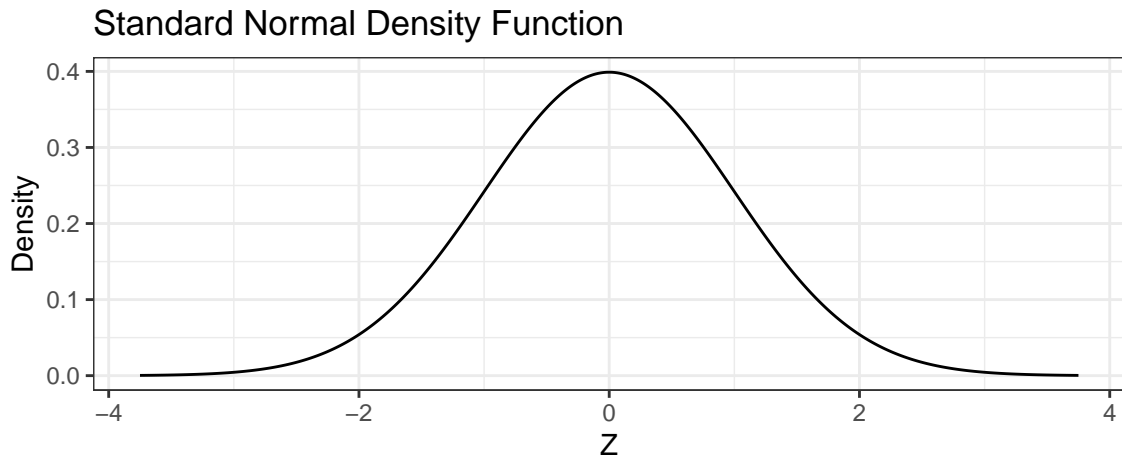
If we assume the forecast errors follow a Normal Distribution, an approximate  $100(1 - \alpha)$  prediction interval can be computed as follows:

$$\hat{F}_t \pm Z * RMSE, \quad (18)$$

where:

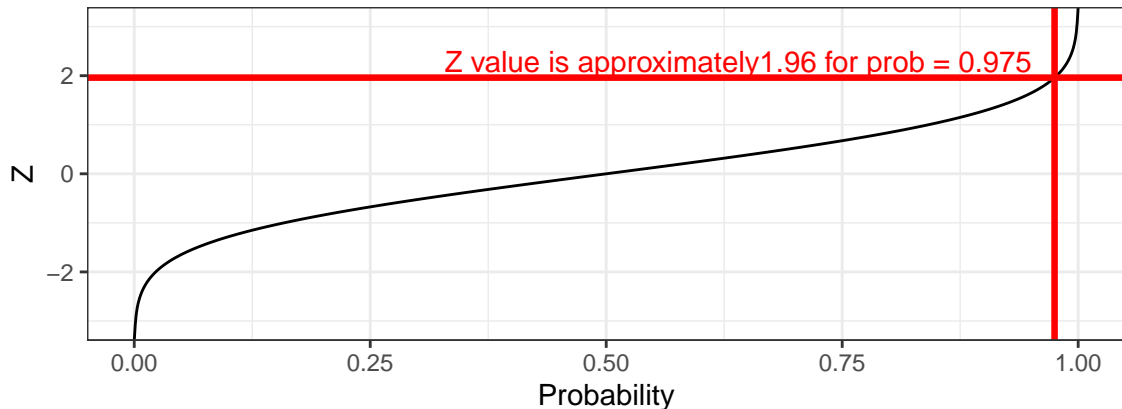
- $\hat{F}_t$  forecast at time  $t$ .
- The RMSE can be used as an estimate of the standard deviation of the forecast errors.
- $Z$  is the quantile corresponding to  $100(1 - \frac{\alpha}{2})$ .

## Recall: Standard Normal Distribution [1]



## Recall: Standard Normal Distribution [2]

Quantile / inverse CDF of Standard Normal Dist.



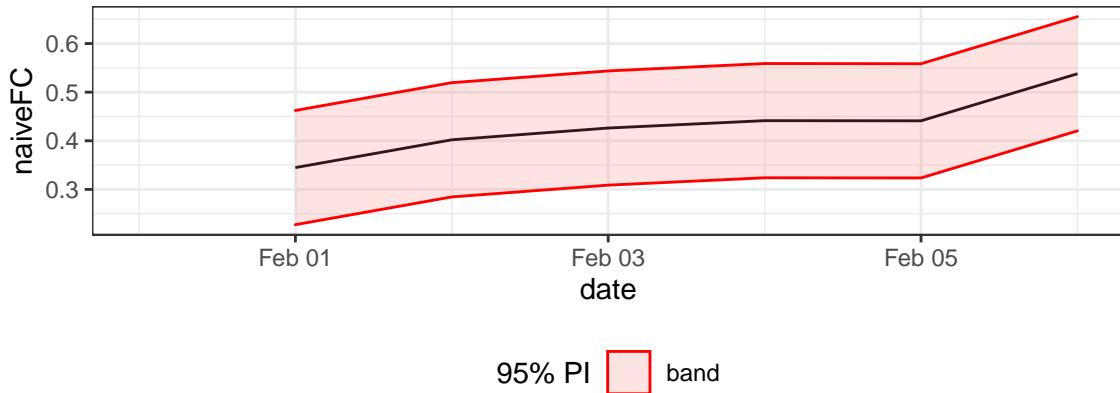


# Prediction Intervals for the \$ADA Data [1]

```
naiveFC = cardano$naiveFC
PInormU = naiveFC + abs(qnorm(0.975))*RMSE
PInormL = naiveFC - abs(qnorm(0.975))*RMSE
dfNaive = data.frame(date = cardano$date, Ft = naiveFC, PInormU, PInormL)
dfNaive %>% ggplot(aes(x= date, y = naiveFC)) + geom_line() +
  theme_bw() + labs(title = "Overlaying 95% PIs in ggplot", fill = "95% PI") +
  geom_ribbon(aes(ymin = PInormL , ymax = PInormU, fill = "band"),
    alpha = 0.2, color = "red") + theme(legend.position= "bottom")
```

## Prediction Intervals for the \$ADA Data [2]

Overlaying 95% PIs in ggplot



# Outline

- 1 Preface
- 2 Transformations
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals
- 5 Recap**

# Summary of Main Points

## Main Learning Outcomes

- Apply transformations to a time series.
- Apply and interpret measures of forecast accuracy.
- Interpret prediction intervals for a simple forecast.

# Things to Do

- **Recommended:** Thoroughly read Chapter 2 of our book.
- Go through the slides, examples and make sure you have a good understanding of what we have covered.
- **Potential Practice Problems:**
  - Extract data using either the `tq_get()` ([tidyquant package](#)) or the `covid19()` ([COVID19 package](#)), and compute the transformations using a manual (i.e., Excel) approach and R. Then, interpret the obtained transformed series.
  - Recreate the charts in Slides 39-41 in [04-Basic-Tools-for-Time-Series-Analysis.pdf](#)
  - **Textbook Example:** For the Means approaches in Example 2.7 (P.49), use R to compute the 7 error forecasting metrics (data available [here](#)).
  - **Textbook Exercise 2.12:** Compute the forecast errors for the naive forecast.
- **Required:** Complete the [graded assignment](#).

## Graded Assignment 05: Evaluating your Understanding

Please go to [Canvas \(click here\)](#) and answer the five questions. **Due February 11, 2021 [11:59 PM, Ohio local time]**.

**What/Why/Prep?** The purpose of this assignment is to evaluate your understanding and retention of the material covered up to the end of Class 05. To reinforce your understanding of the covered material, I also suggest reading all of Chapter 2 of the book.

### General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment.
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the [Remote Desktop](#))

# ISA 444: Business Forecasting

## 05 - Basic Tools for Time Series Analysis

Fadel M. Megahed

Associate Professor  
Department of Information Systems and Analytics  
Farmer School of Business  
Miami University  
Email: [fmegahed@miamioh.edu](mailto:fmegahed@miamioh.edu)  
Office Hours: [Click here to schedule an appointment](#)

Spring 2021