ISA 444: Business Forecasting

02 - Introduction to Time Series Analysis

Fadel M. Megahed

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

Spring 2021

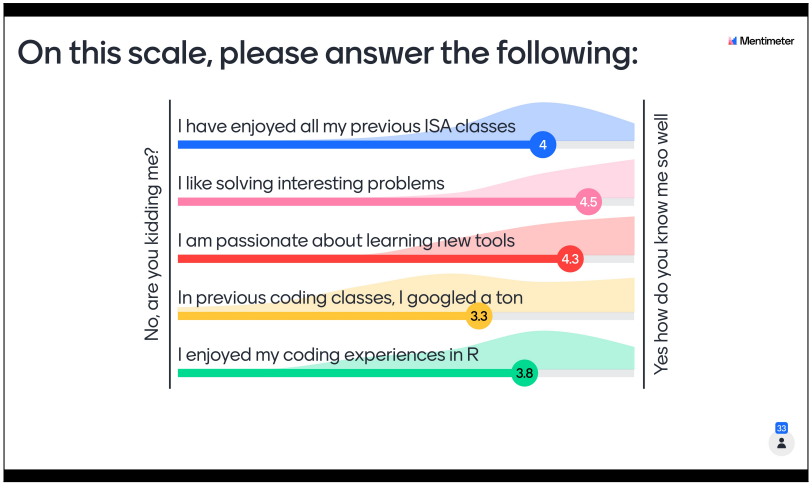
Outline

Quick Refresher based on Last Class

Main Learning Outcomes

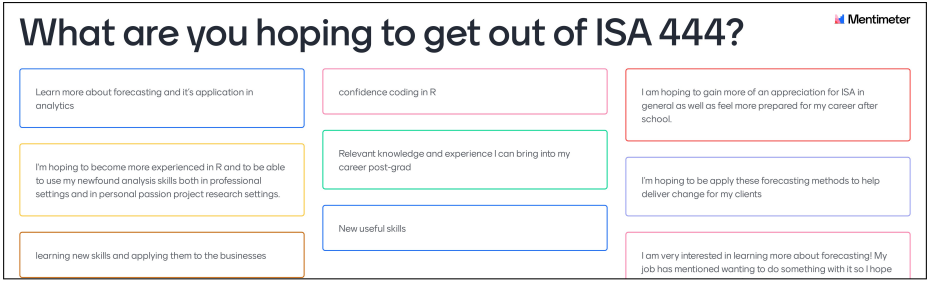
- ☑ **Describe course objectives & structure.**
- ☑ **Describe what do we mean by forecasting**
- ☒ **Explain the PIVASE framework**

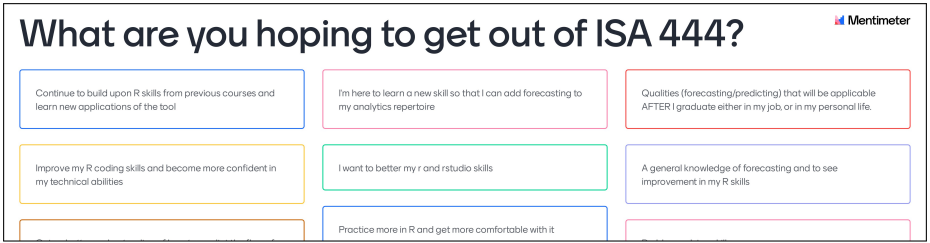
Self-Assessed Skills from Last Class's Mentimeter Poll



Results from Last Class's Mentimeter Poll for Question A

Self-Described Goals for Taking the Class





Learning Objectives for Today's Class

Main Learning Outcomes

- **Explain the PIVASE framework.**
- **Explain the differences between cross sectional, time series, and panel datasets.**
- **Identify and describe the basic components of a time series including trends, seasonal components, and cycles.**

Outline

The COVID-19 Activity

Based on our discussion last class, I have asked to use the COVID-19 package to plot the confirmed COVID-19 cases for the states of Ohio and Kentucky. You should have created two plots: (a) plot containing both of the states as two seperate lines in a line plot, and (b) a paneled plot similar to the output depicted in Slide 21 in Slide Deck 01. **I will quickly walk you through the solution to help ensure that the concepts discussed in last class are well understood.**

Hints:

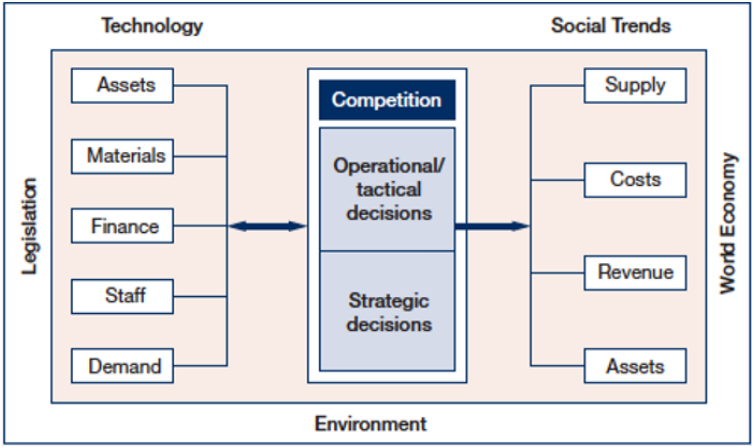
1. Once you have installed the package inspect the `covid19` function using the command `?covid19()`.
2. Set the country argument to the US.
3. Use an apporpriate value for the level argument based on `?covid19()`.
4. Set the starting date to '2020-03-01'
5. Filter the obtained data to only include data for the two states.

Why do we Forecast? – A Conceptual Framework (PIVASE)[1]

- **Purpose:** What do we hope to achieve by generating the forecast? That is, what plans are dependent upon the results of the forecasting exercise? How far ahead do we wish to forecast? We refer to this period as the forecasting horizon.
- **Information:** What do we know that may help us in forecasting. And when will we know it? Detailed data is only useful if it is available in timely fashion.
- **Value:** How valuable is the forecast? What would you pay for perfect knowledge?
- **Analysis:** From analyzing the data can we develop a model that captures its characteristics? And how does it perform on new (hold-out sample) data?
- **System:** What models and software are needed to meet the needs of the organization?
- **Evaluation:** How do we know whether a particular forecasting exercise was effective and what the potential is for improvement?

[0]From: Ord, K., Fildes, R., & Kourentzes, N. (2017). Principles of Business Forecasting (2nd ed., p. 3-6).

Why do Businesses Forecast?



Some of the typical forecasting needs of many organizations.[2]

[2]From: Ord, K., Fildes, R., & Kourentzes, N. (2017). Principles of Business Forecasting (2nd ed., p. 7).

Outline

Cross Sectional Data [1]

Cross Sectional Data: Measurements on multiple units, recorded in a single time period.

Example 1: H1B 2020 Data for Senior Data Scientists at Netflix[3]

EMPLOYER	JOB TITLE	BASE SALARY	LOCATION
NETFLIX INC	SENIOR DATA SCIENTIST	375,000	LOS GATOS, CA
NETFLIX INC	SENIOR DATA SCIENTIST	410,000	LOS GATOS, CA
NETFLIX INC	SENIOR DATA SCIENTIST	450,000	LOS GATOS, CA
NETFLIX INC	SENIOR DATA SCIENTIST	600,000	LOS GATOS, CA

[3]Data scraped from https://h1bdata.info/index.php?em=NETFLIX+INC&job=SENIOR+DATA+SCIENTIST&city=LOS+GATOS&year=2020 on January 26, 2021 using the rvest package in R. The printing was limited to those individuals who started on/after January 01, 2020, with the filters specified in the URL.

Cross Sectional Data [2]

Cross Sectional Data: Measurements on multiple units, recorded in a single time period.

Example 2: Heart Disease Dataset[4]

age	sex	restingBP	maxHR	label
60.00	1.00	130.00	132.00	2
63.00	0.00	108.00	169.00	2
59.00	1.00	178.00	145.00	1
57.00	1.00	152.00	88.00	2
60.00	0.00	158.00	161.00	2
52.00	1.00	125.00	168.00	2
45.00	1.00	128.00	170.00	1
51.00	1.00	140.00	122.00	2
58.00	1.00	140.00	165.00	1
51.00	1.00	100.00	143.00	1
65.00	0.00	160.00	151.00	1
57.00	0.00	120.00	163.00	1
66.00	0.00	178.00	165.00	2

[4]Data sampled from this UCI Machine Learning Repository.

Cross Sectional Data [3]

Cross Sectional Data: Measurements on multiple units, recorded in a single time period.

Example 3: NBA 2020-2021 Leaders - Top 12 in PTS/Game[5]

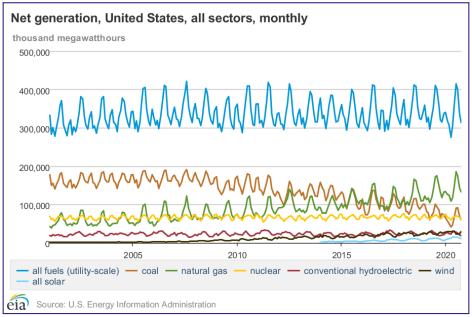
Player	Pos	Age	Tm	G	FG	FG%	eFG%	PTS
Bradley Beal	SG	27	WAS	11	12.2	.489	.538	34.50
Kevin Durant	PF	32	BRK	14	10.1	.518	.593	30.40
Damian Lillard	PG	30	POR	16	8.6	.447	.545	28.70
Stephen Curry	PG	32	GSW	17	9.2	.458	.566	28.40
Kyrie Irving	PG	28	BRK	11	10.5	.511	.593	28.10
Joel Embiid	C	26	PHI	14	8.8	.554	.592	27.70
Jaylen Brown	SG	24	BOS	16	10.4	.528	.595	27.30
Luka Doncic	PG	21	DAL	16	9.6	.458	.507	27.30
Giannis Antetokounmpo	PF	26	MIL	15	9.9	.546	.584	27.00
Zach LaVine	SG	25	CHI	17	9.3	.502	.594	27.00
CJ McCollum	SG	29	POR	13	9.5	.473	.594	26.70
Jayson Tatum	SF	22	BOS	11	10.1	.474	.556	26.60

[5]Data scraped from Basketball-Reference on January 26, 2021 using the rvest package in R. The printing with limited to the top 12 players and the selected variables.

Time Series Data [1]

Time Series Data: Comparable measurements recorded on a single (or a few) variables over time (usually a long period of time).

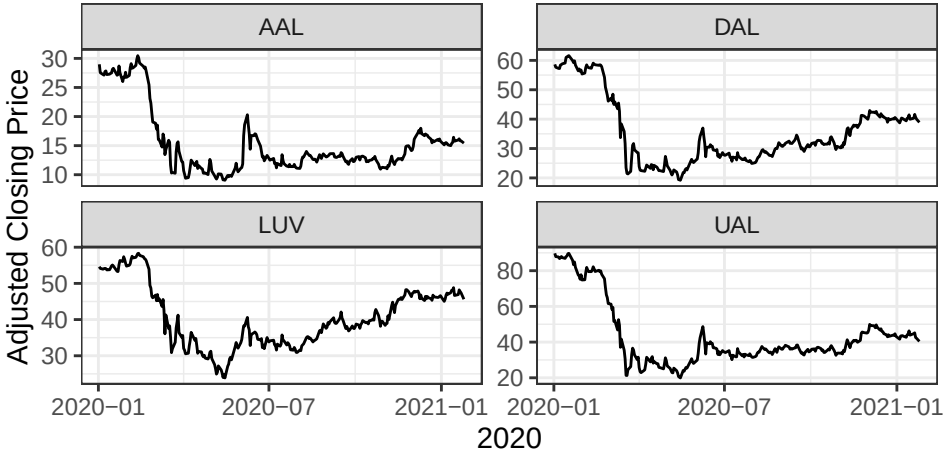
Example 1: Net Power Generation in the U.S.



Net monthly electricity generation in the U.S. from the U.S. Energy Information Adminstration.

Time Series Data [2]

Example 2: Stock prices of U.S. Airlines



As of January, 26 2021: COVID-19 had a substantial impact on these airline stocks.

Panel Data

Panel Data: Cross sectional measurements (usually many variables) repeated over time (usually over a few time periods).

Example: World Bank's Data[6]

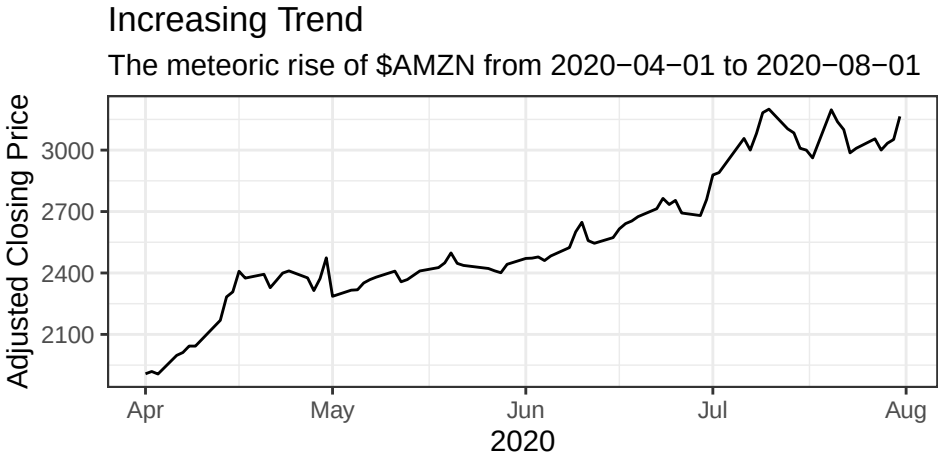
iso3c	date	NY.GDP.MKTP.KD.ZG	SH.DYN.NMRT	SH.HIV.INCD.ZS	SH.MED.BEDS.ZS	SH.MED.PHYS.ZS	SH
CHN	2018.00	6.75	4.20		4.31	1.98	
CHN	2019.00	6.11	3.90		4.31	1.98	
CHN	2020.00	6.11	3.90		4.31	1.98	
EGY	2018.00	5.31	11.50	0.08	1.43	0.45	
EGY	2019.00	5.56	11.10	0.09	1.43	0.45	
EGY	2020.00	5.56	11.10	0.09	1.43	0.45	
USA	2018.00	2.93	3.80	0.22	2.87	2.61	
USA	2019.00	2.16	3.70	0.22	2.87	2.61	
USA	2020.00	2.16	3.70	0.22	2.87	2.61	

[6] Data queried from the World Bank Data Catalog using the wbstats package in R. The printed results show a snapshot of 7 variables (out of a much larger panel dataset). You can think of panel data as a cross-sectional dataset with a longitudinal/time component.

Outline

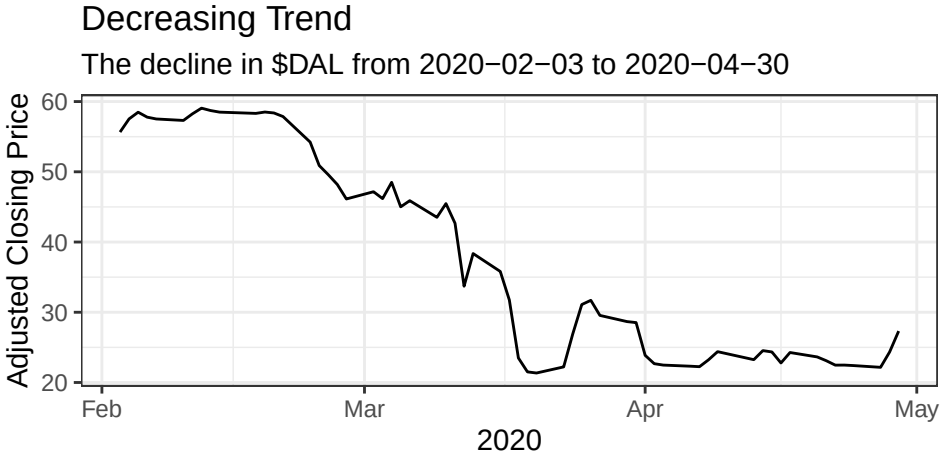
Trend [1]

A **trend** is an increasing or decreasing pattern over time.



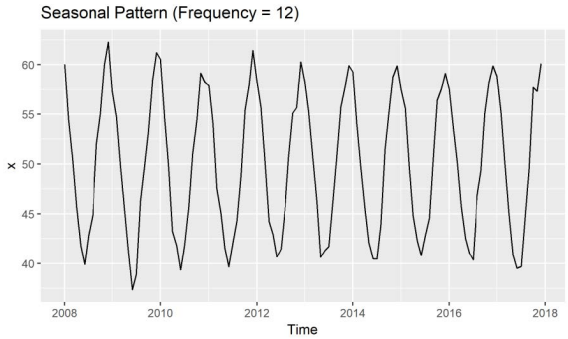
Trend [2]

A **trend** is an increasing or decreasing pattern over time.



Seasonality [1]

Seasonality refers to the property of a time series that displays REGULAR patterns that repeat at a constant frequency (m).

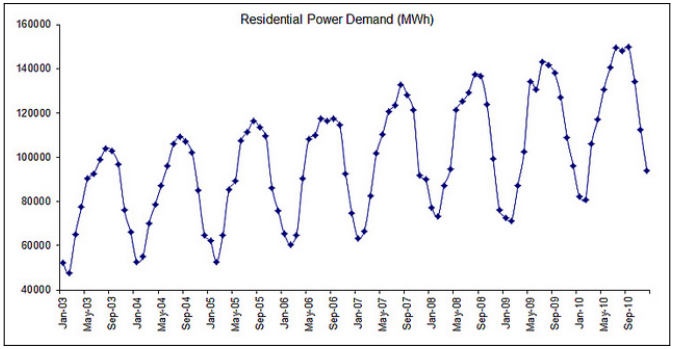


A time series with a monthly seasonal pattern.[7]

[7] Figure is from Dr. Allison Jones-Farmer's lecture notes, Miami University, Spring 2020.

Seasonality [2]

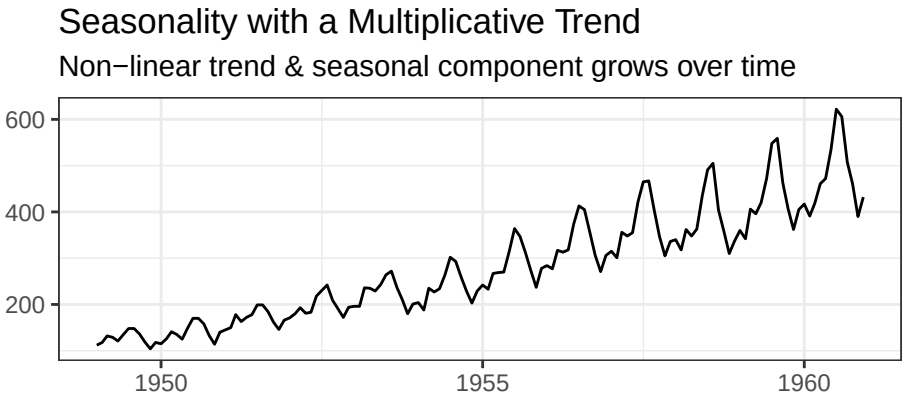
Seasonality refers to the property of a time series that displays REGULAR patterns that repeat at a constant frequency (m).



Seasonality with an Additive Trend[8].

[8]Figure is from Patterns Unplugged - Help Center published online on December 28, 2016.

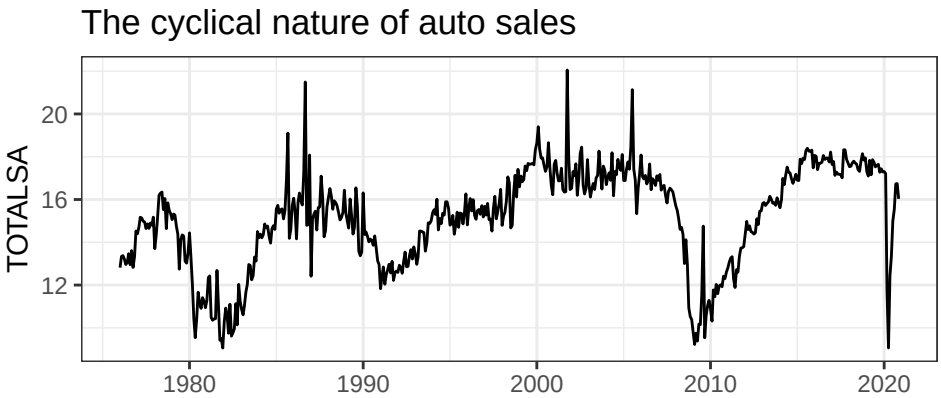
Seasonality [3]



AirPassengers R Dataset −− Source: Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1976) Time Series Analysis, Forecasting and Control.

Cycle [1]

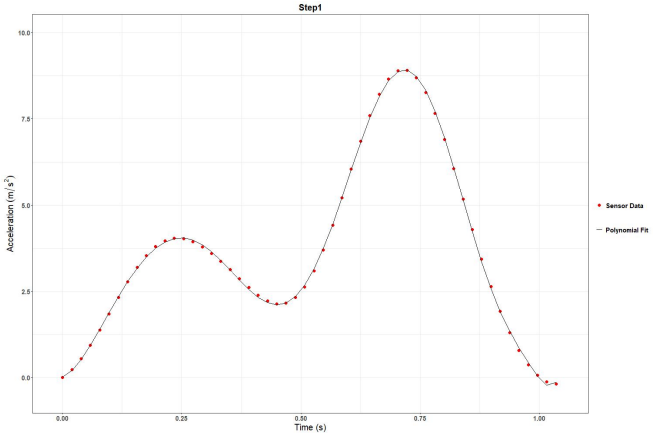
Cyclical fluctuations are somewhat irregular (unknown duration).



Total Vehicle Sales [TOTALSA], retrieved from FRED, Federal Reserve Bank of St. Louis https://fred.stlouisfed.org/series/TOTALSA, January 26, 2021.

Cycle [2]

Cyclical fluctuations are somewhat irregular (unknown duration).



A gait "cycle" from an IMU attached to the ankle. Joint work with the University at Buffalo.

Outline

Summary of Main Points

Main Learning Outcomes

- **Explain the differences between cross sectional, time series, and panel datasets.**
- **Identify and describe the basic components of a time series including trends, seasonal components, and cycles.**

Things to Do

- Thoroughly read Chapter 1 of our book, which can be downloaded from the Publisher (if you have not gotten your book yet).
- Go through the slides, examples and make sure you have a good understanding of what we have covered.
- **Optional, but highly recommended,** go through the self-paced interactive review for Week 01.
- **Required:** Complete the graded assignment.

Self-Paced Interactive Review for Week 01

To help you brush your R skills and assess your understanding of the material covered in Week 01, I have created a self-paced interactive review. The review can be accessed at http://rstudio.fsb.miamioh.edu:3838/megahefm/isa444/week01/.

This review is optional, but if you plan on completing it, you are advised to do the following prior to going through the self-paced review.

1. Read Chapter 01 from our textbook.
2. Have completed the in-class examples in the lecture notes for Week 01.
3. Be able to dedicate about 20-40 minutes to go through this entire tutorial.

You are encouraged to complete this review prior to our first Week 02 class; there is no submission associated with this review.

Graded Assignment 02: Evaluating your Understanding

Please go to Canvas (click here) and answer the four questions. **Due February 01, 2021 [11:40 AM, 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 02. In order to prepare for this, you should have either actively attended class and/or watched the recording from WebEx. Furthermore, you should have thoroughly read Chapter 01 of the book.

General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment. If the concepts we covered are well-understood, this should take 20-35 minutes.
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the Remote Desktop)

ISA 444: Business Forecasting

02 - Introduction to Time Series Analysis

Fadel M. Megahed

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

Spring 2021