Module 1: Introduction to Financial Analytics and Time Series Data

Table of Contents

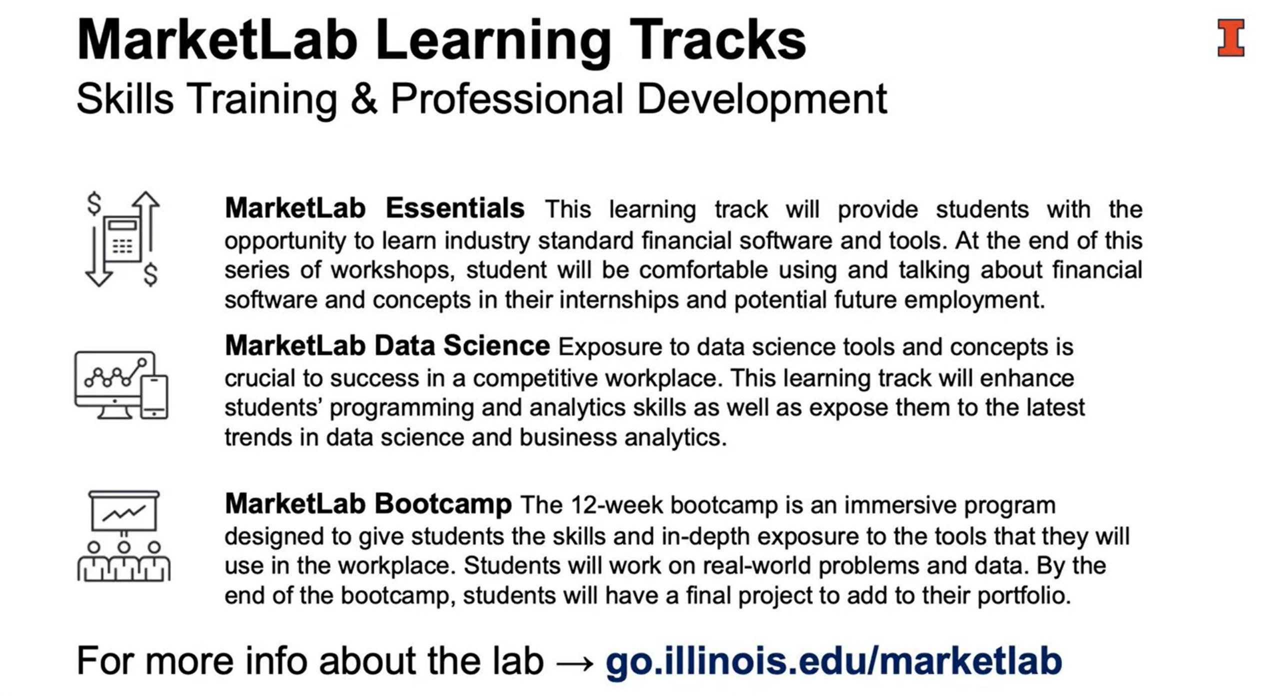
About the Course

[Coursera Course Introduction]

Hi, and welcome to selected topics in Financial Analytics. In this class and in the next set of videos, we're going to apply different analytic techniques to finance and we'll be using a number of different data sources, with an emphasis on time series data. What do we mean by time series data? Think about the number of pulley transit riders every day, week, and month. If we record this type of data over time, we will have a time series data set. So, when we look at stock prices, we'll be looking at different time periods for an hourly to daily, monthly to quarterly, and annual prices. We will analyze the behavior and changes over a set period of time. How do we analyze time series data? There are various analytic techniques such as moving averages, exponential smoothing, and ARIMA models, which we'll be using to forecast time series data. After working with these analytic techniques, we will look at a stock prices from a portfolio perspective. A portfolio can be created by grouping different stocks with common or uncommon characteristics. We will then use portfolio analytic techniques to evaluate the risk and rewards and expected returns of that group of stocks. By the end of this class, you should have a good understanding of the time series data, different analytic techniques, and how to evaluate a portfolio of stocks. In the next four weeks we'll be working with a specific cases, but hopefully after this class, you will be able to dive deeper into some of these topics and applications. Welcome, and I hope that you enjoy this class.

Instructor Bio: Jose Rodriguez

Hello, I'm Jose Luis Rodriguez. I currently serve as the director of the Margolis market information lab. And I'm also an RC Evans innovation fellow. Part of the disruption lab, at Gies College of Business.



MarketLab Learning Tracks

Skills Training & Professional Development

MarketLab Essentials This learning track will provide students with the opportunity to learn industry standard financial software and tools. At the end of this series of workshops, student will be comfortable using and talking about financial software and concepts in their internships and potential future employment.

MarketLab Data Science Exposure to data science tools and concepts is crucial to success in a competitive workplace. This learning track will enhance students' programming and analytics skills as well as expose them to the latest trends in data science and business analytics.

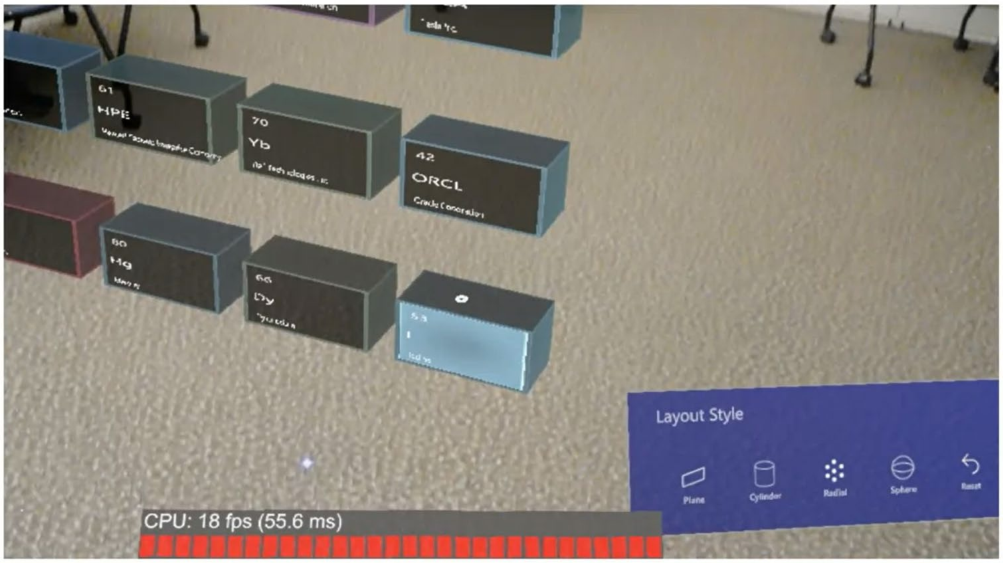
MarketLab Bootcamp The 12-week bootcamp is an immersive program designed to give students the skills and in-depth exposure to the tools that they will use in the workplace. Students will work on real-world problems and data. By the end of the bootcamp, students will have a final project to add to their portfolio.

For more info about the lab → **go.illinois.edu/marketlab**

At the Margolis market information lab, we offer a number of learning tracks covering and a range of financial software. We also offer data science crash courses in Python, R and Tableau. On 12-week boot camps where students are exposed to finance and analytics tools used in the job market. In addition to leading the market lap, I teach quantitative and data science courses for the Master of Finance and the IMBA graduate programs. I'm also involved in a number of research projects on campus.

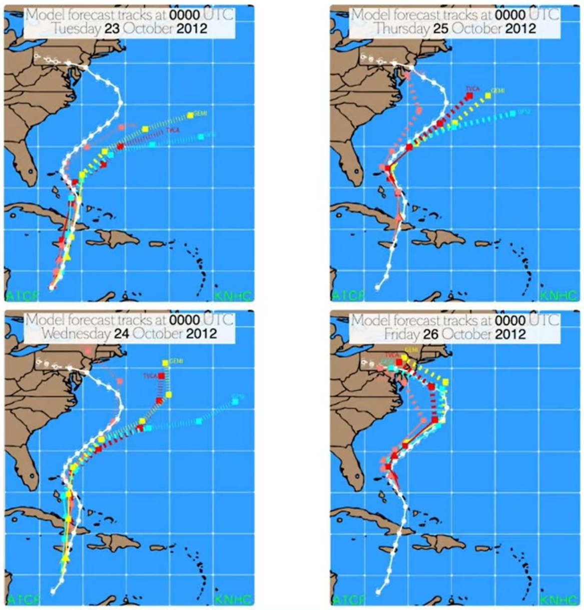
Augmented Reality App Prototype

Finance Dashboard



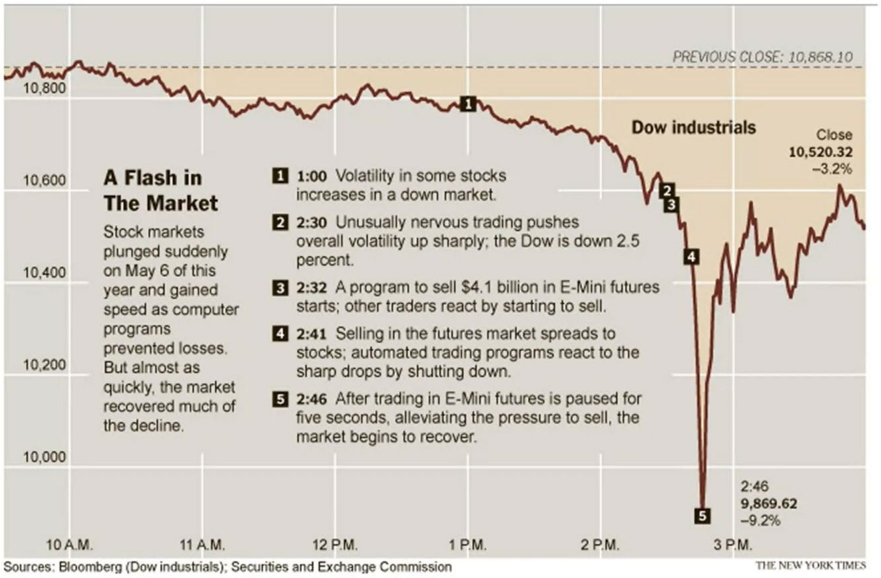
One very exciting project I'm working on is that, Augmented Reality Application on HoloLens 2, to build, visualize and interact with the market. Prior to joining the finance department at this college of business, I serve as assistant director and data scientists at the CME business analytics lab. And taught business analytic courses. I quill in the school of business, at Loyola University of Chicago. A home, on the graduate degrees in mathematics and computer Science. As well as master's degrees in computer science from Loyola University, Chicago.

Research Interest


Model forecast tracks at 0000 UTC
Tuesday 23 October 2012
Model forecast tracks at 0000 UTC
Thursday 25 October 2012
Model forecast tracks at 0000 UTC
Wednesday 24 October 2012
Model forecast tracks at 0000 UTC
Friday 26 October 2012
ATCF
KNHC

I enjoy reading and doing research on how natural disasters impact the stock market. For example, Hurricane Sandy impact on the stock market.

Research Interest



A Flash in The Market

Stock markets plunged suddenly on May 6 of this year and gained speed as computer programs prevented losses. But almost as quickly, the market recovered much of the decline.

1 **1:00** Volatility in some stocks increases in a down market.

2 **2:30** Unusually nervous trading pushes overall volatility up sharply; the Dow is down 2.5 percent.

3 **2:32** A program to sell $4.1 billion in E-Mini futures starts; other traders react by starting to sell.

4 **2:41** Selling in the futures market spreads to stocks; automated trading programs react to the sharp drops by shutting down.

5 **2:46** After trading in E-Mini futures is paused for five seconds, alleviating the pressure to sell, the market begins to recover.

Sources: Bloomberg (Dow industrials); Securities and Exchange Commission

Additional Past research include the 2010 market crash and low latency transactions on high frequency in financial markets. I have developed a number of software packages in different areas such as an analysis of the United States pivens and trademark data. News on areas data from seeking alpha.com. You can find some of these open source packages on my GitHub. Prior to joining the Gies College of Business, I lived in Chicago for about six years, and before that, I lived in New York City for five years.

Ragnar Relay Series

200 Miles Race

You can usually catch me running a 200 mile races over complicated in my house with unnecessary electronics.

Finance & Data Science Teaching

I'm working with a students at the Margolis markets inflammation lab.

Interview with Jose Rodiguez

In the first part of this class, we will be exploring some elements of financial analytics. We have the pleasure today of talking to Jose Rodriguez who's the director of the Margolis Market Information Lab, here at the University of Illinois, East College of Business. He works a lot with financial data. So I'm hoping that he can share with us some of his insights about financial analytics and how to apply them in a real-world setting. So Jose, I wan you to introduce yourself for the audience here.

Thank you Son. I'm the Director of the Margolis Market Information Lab. My main role of what we do here at the lab is we have access to different softwares and datasets from high frequency to unroll. We train students and we give access to faculty and train students on how to you use this dataset and prepare them for the workplace. We have BookCons, we have different ways to train our students to be prepared when they go out there to the workplace to get ready to any challenge that there may be tossed. Maybe you could tell us a little bit about your background before you arrived here at the University of Illinois and some of your past experience. So my background is math and computer science, and I have graduate degree in computer science. I worked before coming here at of Illinois at Loyola University of Chicago, at the School of Business there. I did for about four years of research on work with the CME Group on their high-frequency data. What is the CME Group? The CME Group is a market-maker of marketplace for futures and options. So they trade between commodities, index, equities, and all products that they have.

What does CME stands for? CME, Chicago Mercantile Exchange. The Chicago Mercantile Exchange. What do you love about the world of finance and financial data? With my background and how I came to finance. I loved a blast datasets and how complex the data is, the type of problems that we are tasked with, the diverse permanence from agriculture to technology, to the environment, to any datasets that we can apply, I'm bored with. I've enjoyed that a lot. At the time, working with all the new analytic tools at the new analytic techniques that are all there applied to these datasets to gain insight. So do you have any advice for students who are preparing for a career in financial analytics? I'll say to know their backgrounds, so they should know math, the analytics. I know that financial theory will board whatever job that are getting to.

So broadly speaking, can you tell me a little bit about the role of analytics in today's marketplace or today's trading companies or financial companies? I would say that with the increase of the data, that's what analytics have gotten so popular in new analytics techniques. Computers have gone too very powerful. So knowing and applying different analytic techniques, I'm being able to work with the equipment, computers from coding to applying to theory, to financial theory, at the same time, is very valuable. Yeah. That does make sense. Then broadly speaking in the financial world, how do firms collect data? Do they collect it by themselves or do they purchase that data? Where does the data come from? Is there a preprocess that firms have to do with the data before they can actually begin the analytic process/ So usually, what I will say depending on the firm or depending where you go, there always will be there part of compliance.

So you collect everything you do, you keep it to one side. You have your team that make sure that your data is clean, it's organized, is in the right way, and you have your order theme to quantitative research or some research that is maybe tried later for your front desk. So there are different teams. They all work the same dataset, different in piles of data. Depending of where you going in the firm, then you will find what type of data you're working with. So broadly speaking, what are the key things a manager would need to know if they're managing a financial analytics team? I'll say, they need a good team. So they have to start from the ground with the right people. Now, they say we're in talks about data scientists or if you want to have someone that knows this statistics, they have the background in finance, they have the background in encoding, and can also talk with everyone else in their team to deliver the message it faces from the quantitative side, so to explain. You're all the managers that maybe are not that bears on the quantitative side. What are the ins and outs of astrology? Or whether you are looking for, that's very important.

Lesson 1-0: Module 1 Introduction

Module 1 Overview

In this module, we will introduce an overview of financial analytics. Students will learn why, when, and how to apply financial analytics in real world situations. We will explore techniques to analyze time-series data and how to evaluate risk reward trade-off in modern portfolio theory. While most of our focus will be on prices, returns and risk of corporate stocks. The analytic techniques can be leveraged in other domains. Some of the topics that we will be covering will be elements of forecasting, introduction to financial analytics, and performance metrics.

Jose Rodriguez: Forecasting in Practice

Welcome to this week's lecture. I'm back with Jose. This week we'll be talking a little bit about what are the basic financial analytic tools, so maybe Jose can help us out. So Jose, what are some of the common analytic tools that you have seen in the financial analytics workplace? So from the software side, we're very heavy on R and Python, of course. Then for generally analytics tools or analytic tool kit, I recall, will be linear models. So have a good understanding of linear models. Like regression, linear. Regression analysis, logistic regression, all type of regressions, dimensionality reduction, and PCA or Principal Component Analysis being able to produce different features that you may have. Right, and that reduces the complexity and also helps to optimize the algorithms and speed them up. Correct. So if you have a large data-set with many features or many different, columns, if you are able to produce that using PCA or any dimensional IT options, then you will be able to optimize your algorithm as well. Those are the more common ones or the traditional ones and newer or currently will be a narrow networks and three's and deep learning or machine-learning techniques. Those are the more trendy ones that currently everyone sees. Yes, today. I know that financial data is really time-series data. So maybe you could help us understand a little bit about time-series data and the financial analytics context? Time-series data from my point of view when defining what scale you're working is one of the first challenges that you find, if you are working with monthly, daily, weekly, how low do you go in your scale depending on your problem. From my experience and my background, I've worked with data that is less than a minute, so we call it ticket data. I'm going all the way down to seconds, milliseconds, and nanoseconds data. Wow that's fast. All right, so we will begin week one of our course.

this person bought at three gas stations in a row and so no, that sounds like fraud, let's stop it and let's apply a rule. So actually, I wanted you to think of your day so far, I don't know how it's going. But think how many times you use rules in trying to answer a phone call or an email or have lunch. You will see a whole day was probably decomposed into series of decisions which depended on the way you made rules. So once you are very effect, you can say, okay, let's see why these rules came about.

Lesson 1-1: Elements of Forecasting

Lesson 1-1.1 Subjective Forecasting

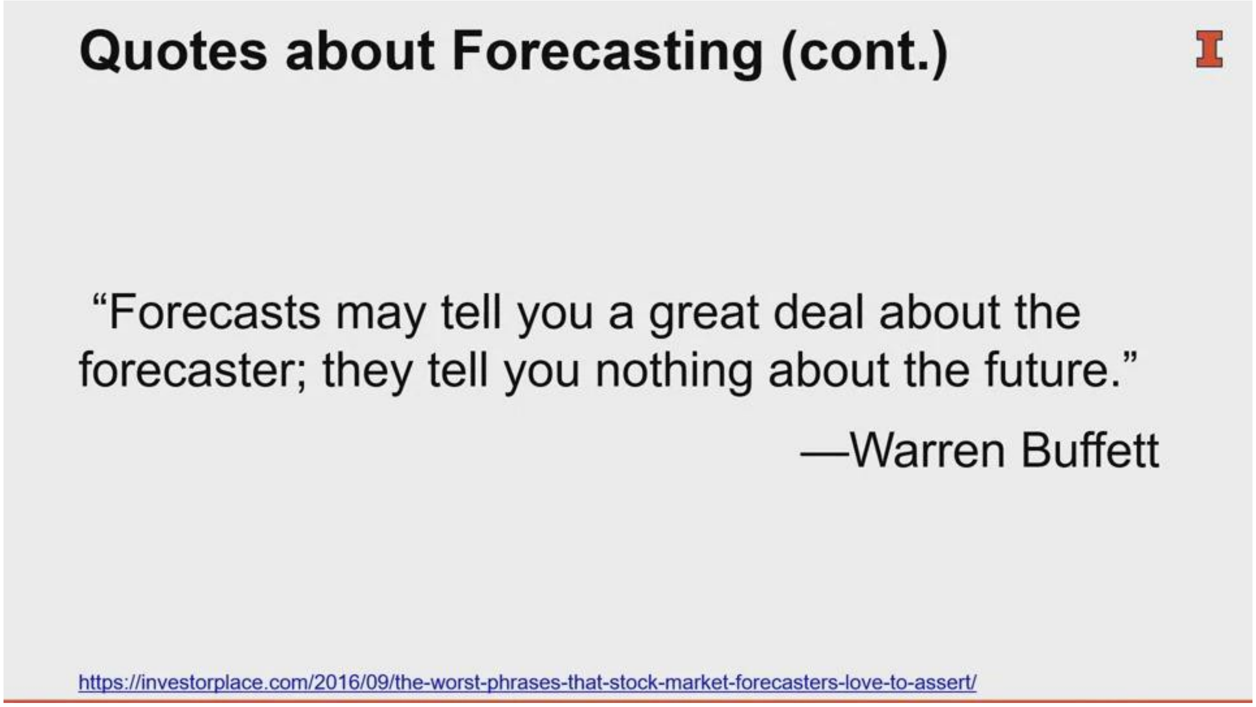
Quotes about Forecasting

"Those who have knowledge, don't predict. Those who predict, don't have knowledge.“

—Lao Tzu, 6th Century BC Chinese Poet

http://www1.secam.ex.ac.uk/famous-forecasting-quotes.dhtml

Hello. In this set of slides, we'll talk a little bit about what is forecasting? Then more specifically, what is forecasting in a business context? So the set of slides will give an introduction to the lectures that will follow. But before we begin, let's look at some quotes about how past people, great thinkers have thought about what forecasting is. Here's a code from Lao Tzu who said, "Those who have knowledge don't predict, and those who predict don't acknowledge." This certainly gets to one of the difficulties of forecasting, and that has to do with uncertainty. We don't really know what's going to happen in the future. No one does. We don't have a crystal ball, and I think that's what Lao Tzu was trying to say here.
you figure it out?" So this person is blue collar I know, they're filled in a form with all their preferences, whether this person is single, has tertiary education, probably a PhD, has never default in a loan, and still rents. Okay. Fine. So one way you say, it's a simple way, look at all the people who have bought the product and see if there is somebody similar to this person who has bought this product.



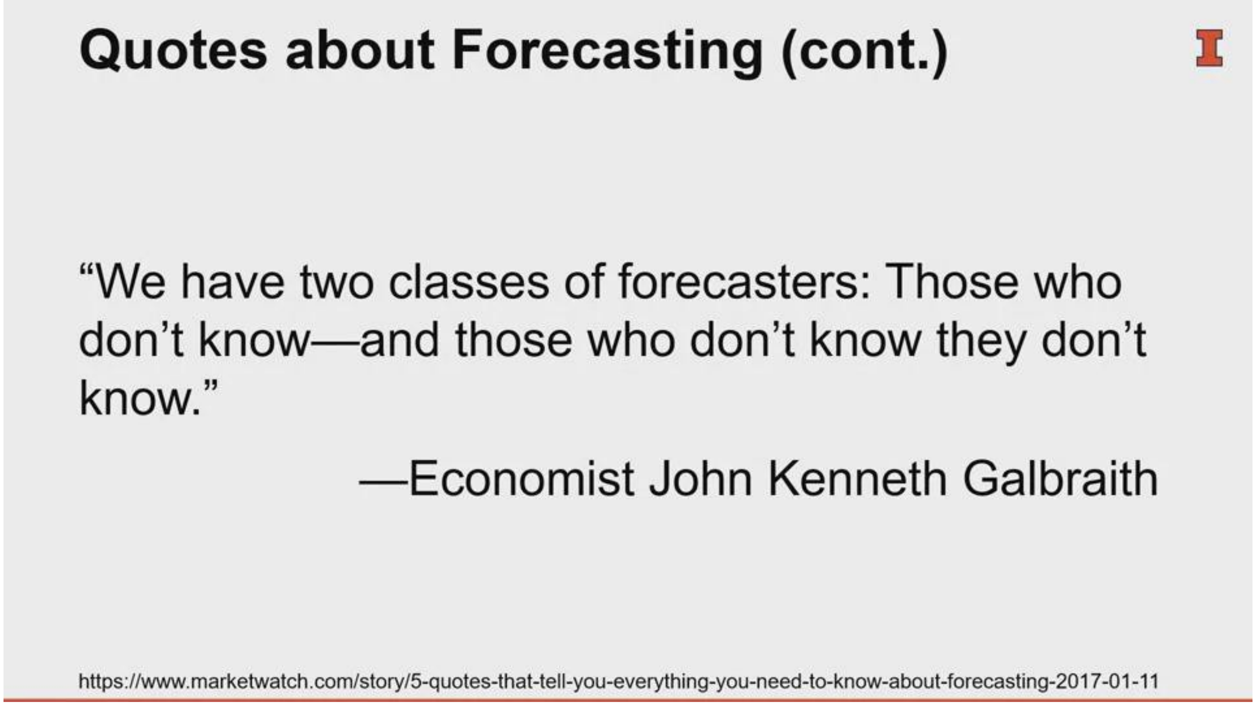
Quotes about Forecasting (cont.)

“Forecasts may tell you a great deal about the forecaster; they tell you nothing about the future.”

—Warren Buffett

https://investorplace.com/2016/09/the-worst-phrases-that-stock-market-forecasters-love-to-assert/

Here's another quote by Warren Buffett who says that forecasts may tell you a great deal about the forecaster, but they tell you nothing about the future. Certainly, Warren Buffett believes this. He doesn't use traditional methods in finance. He actually believes that he tries to restrict himself to a circle of competence as he calls it. So he tries to stay within his domain when making investment decisions. He tries to stay within a business domain. He tries to have a deep understanding of what these businesses are. Notably he does not invest in textbox because he doesn't have a clear understanding of what they're about.



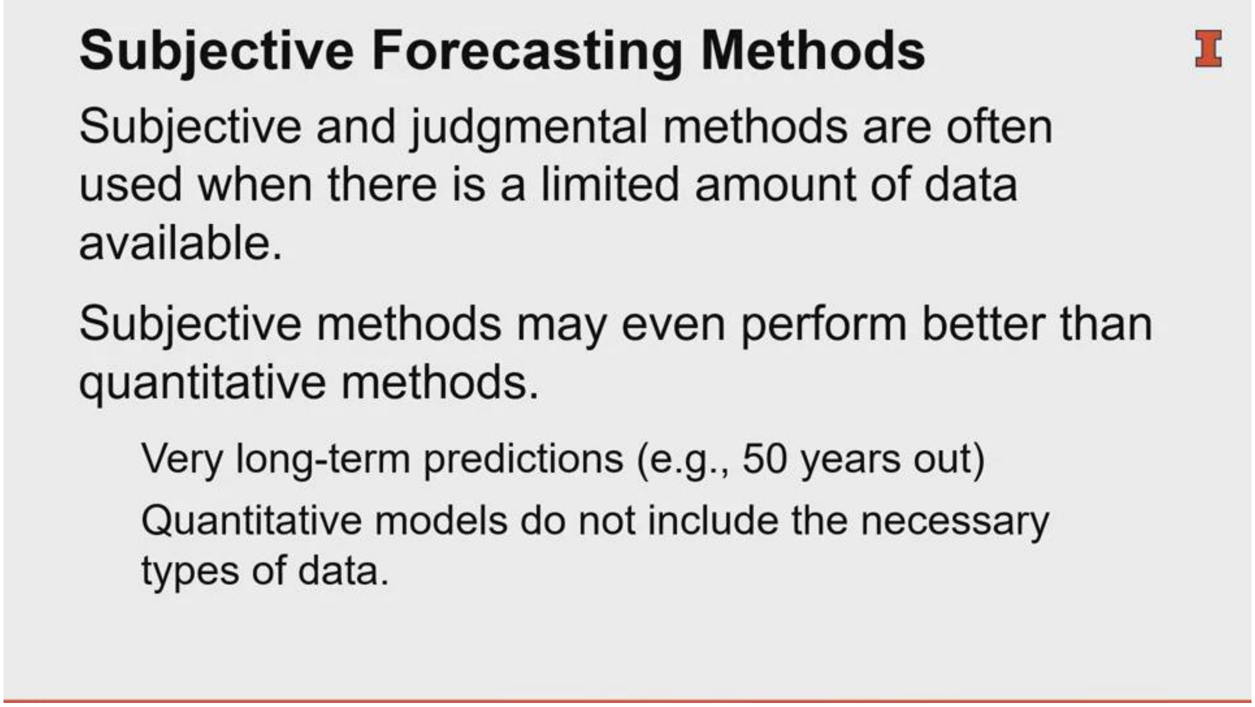
Quotes about Forecasting (cont.)

“We have two classes of forecasters: Those who don’t know—and those who don’t know they don’t know.”

—Economist John Kenneth Galbraith

https://www.marketwatch.com/story/5-quotes-that-tell-you-everything-you-need-to-know-about-forecasting-2017-01-11

Then finally we have this last quote by John Kenneth Galbraith who said there are two types of forecasters, those who don't know, and those who don't know that they don't know. So I think all three of these get to this idea of uncertainty of trying to predict the future.



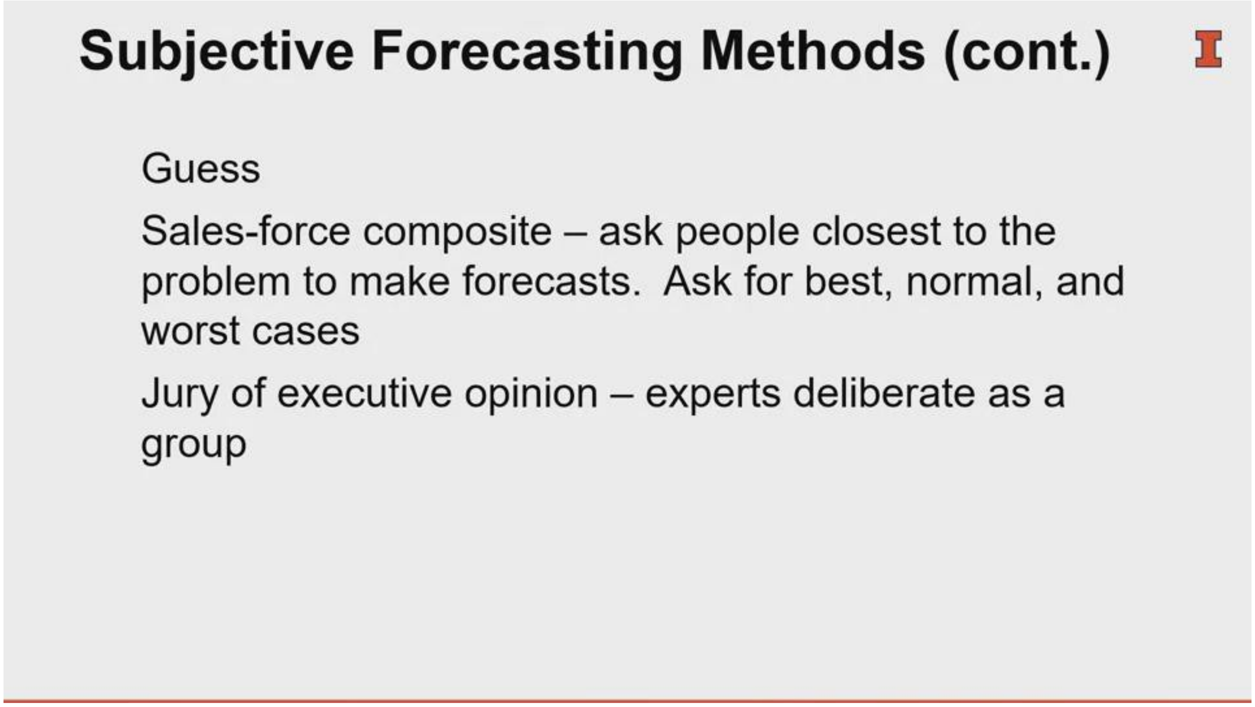
Subjective Forecasting Methods

Subjective and judgmental methods are often used when there is a limited amount of data available.

Subjective methods may even perform better than quantitative methods.

- Very long-term predictions (e.g., 50 years out)
- Quantitative models do not include the necessary types of data.

But nonetheless, there are some techniques that people have tried out in the past. I'm going to talk a little bit about some subjective forecasting methods that have been used. These are judgmental methods that were there. People make decisions based on limited amounts of data, and sometimes they're even better than the quantitative methods that we'll be focusing on in this section. They're sometimes better for long-term predictions, 50 years out, or a 100 years out using just historical data, and then qualitative models also don't have the necessary types of data that might be needed for long-term forecasts. So what are some of these subjective forecasting methods? The first one is we can simply guess. What do you think the weather will be like tomorrow? It's going to rain, it's going to be sunny. That's a guess. It's a wild guess. But clearly, that's not the best method we could use. So people have come up with some subjective forecasting methods.



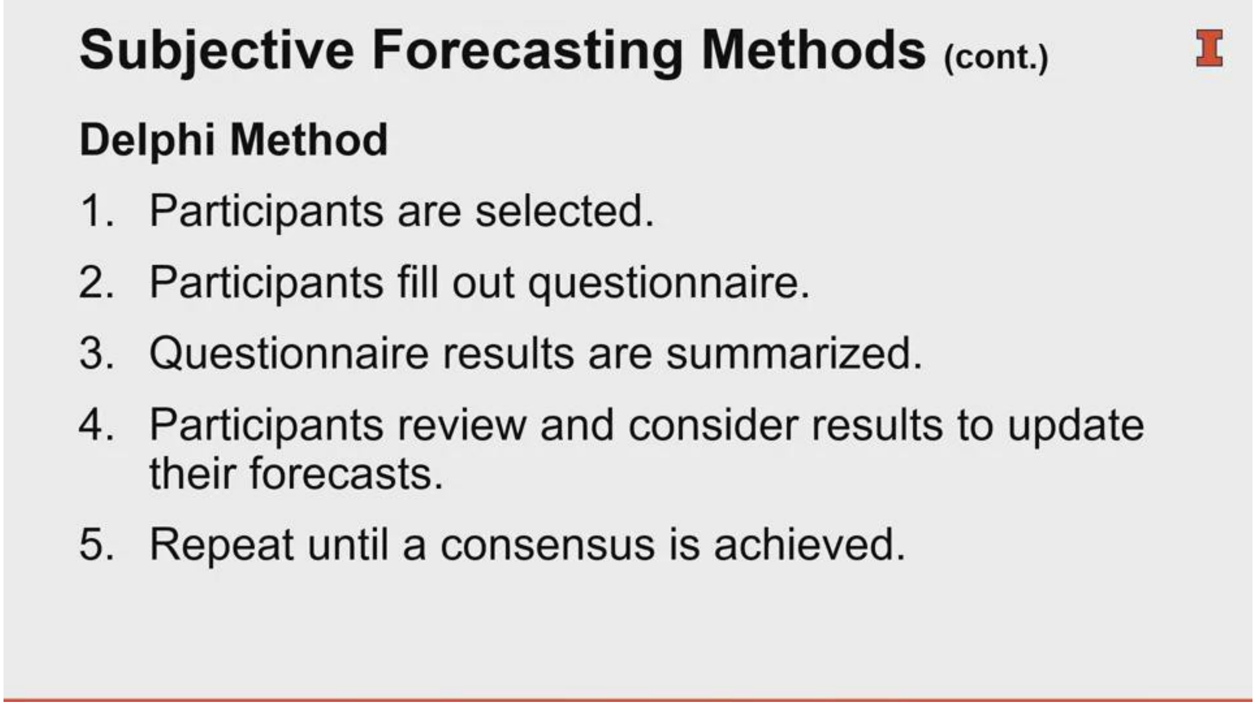
Subjective Forecasting Methods (cont.)

Guess

Sales-force composite – ask people closest to the problem to make forecasts. Ask for best, normal, and worst cases

Jury of executive opinion – experts deliberate as a group

The first one I can describe as the salesforce composite. Basically, the idea here is to ask people closest to the problem to make forecasts based on their best judgment. Sometimes we might ask for what's the best-case scenario? Or what's the worst-case scenario? What's something in between? Another method relate to the salesforce can positive you're trying to predict sales is to ask the customers. What do they think? Do they like the product? Do they not like the product? Would they buy the product? So that's basically the sales force composite. Next, we can have a jury of executive opinion. That's where we bring together a number of executives, experts in a field, medical doctors for example, legal professionals, management consultants to come together and they will deliberate on a set of facts, and as a group, come up with some forecasting decision.

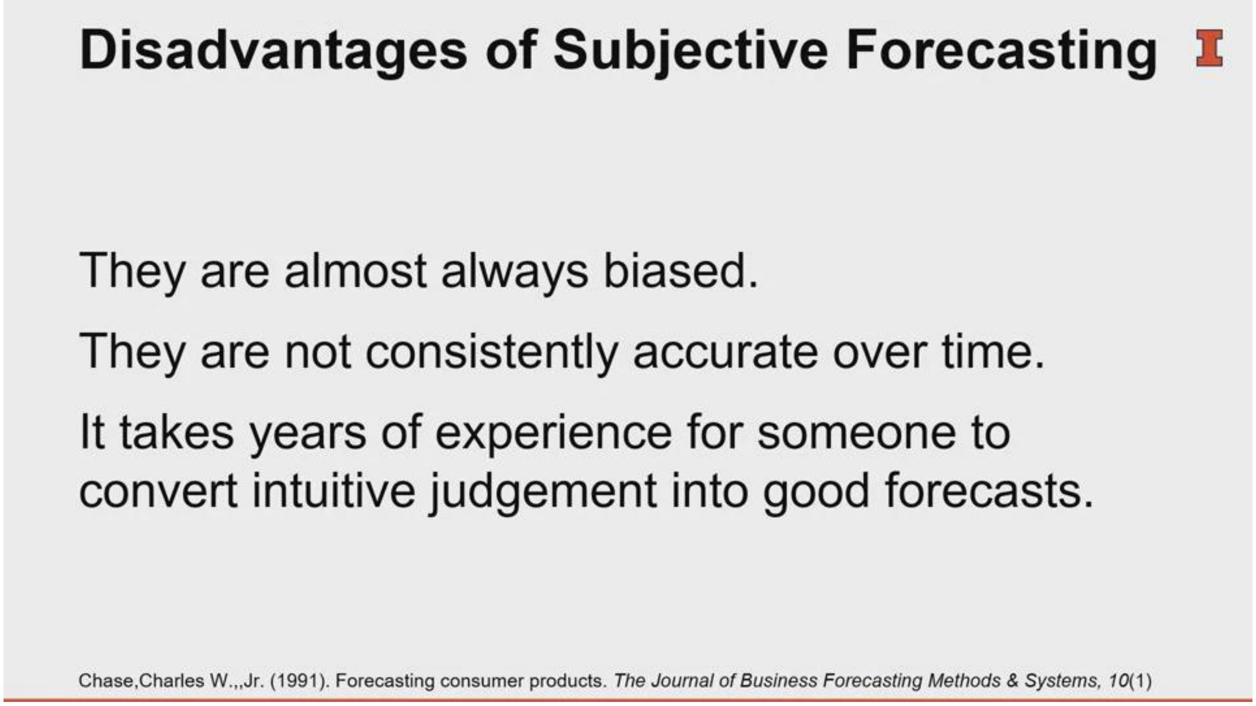


Subjective Forecasting Methods (cont.)

Delphi Method

1. Participants are selected.
2. Participants fill out questionnaire.
3. Questionnaire results are summarized.
4. Participants review and consider results to update their forecasts.
5. Repeat until a consensus is achieved.

The next method is a pretty robust method for subjective forecasting models go. It's known as the Delphi method. It's similar to the jury of experts. But slightly different. We have the first step is to select the number of participants. These are usually experts in their field, or have experience in the subject matter. Then they're asked to fill out a questionnaire. The questionnaire is designed to try to get their understanding of the facts and make a prediction or some forecast. Then a separate team take the questionnaire, and summarize the results. These results are then passed back to the participants for review, and they consider the results which includes the results of very one, and they update the forecast. This process is repeated until there is some consensus achieved among all the participants. Now, this method does have some advantages over he jury of experts. One, it negates the influence of people in a jury who might have a strong influence. They might be very charismatic. They might have a higher rank in an organization, or they might exhibit some bias toward the problem. That could be racial or gender bias or something like that. So by having the participants unknown to each other, they are freer to give more honest answers without the influence of others. So that's the Delphi method.



Disadvantages of Subjective Forecasting

They are almost always biased.

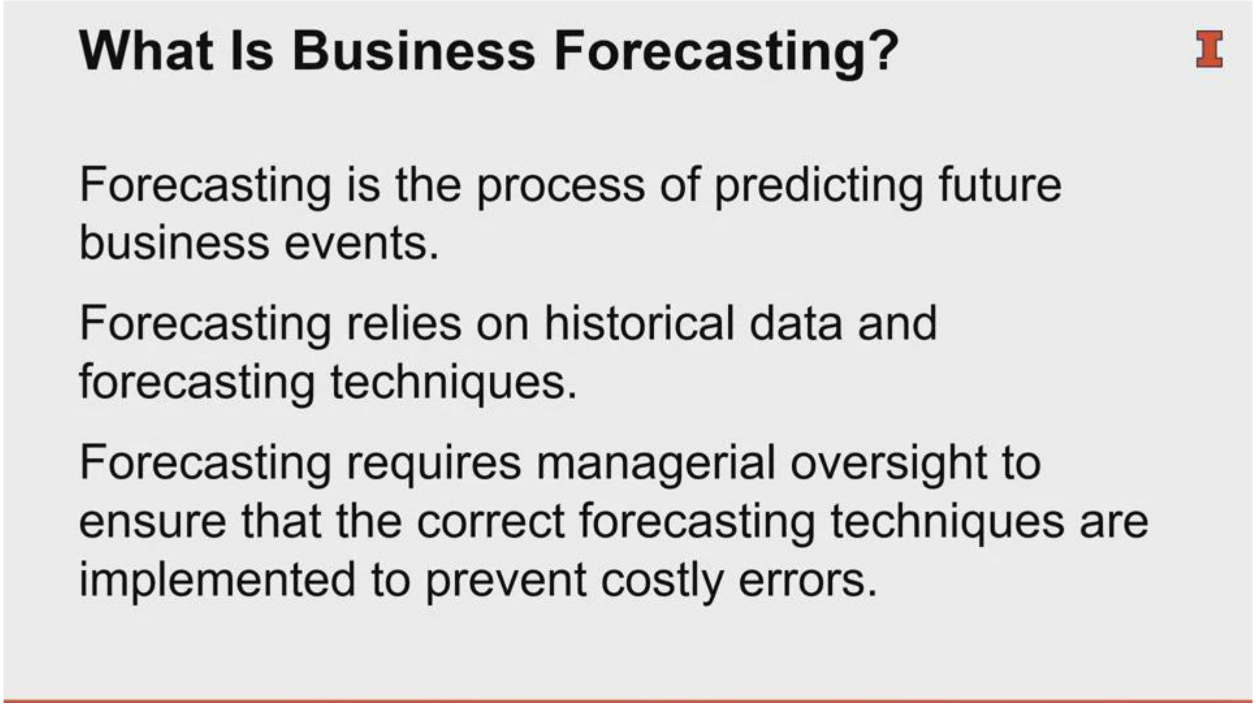
They are not consistently accurate over time.

It takes years of experience for someone to convert intuitive judgement into good forecasts.

Chase,Charles W.,,Jr. (1991). Forecasting consumer products. *The Journal of Business Forecasting Methods & Systems, 10*(1)

But in general, there are some disadvantages of subjective forecasting methods. More likely than not, they're always biased. Biased, meaning they lean one way or the other. Another problem with subjective forecasting is that they're not consistently accurate over time. So we would like to have forecasts that are accurate over time, and also to have it to be a repeatable process. Then finally, to make good subjective forecasting decisions, it takes years of experience for someone to convert intuitive judgment into good forecast. That's why the CEO of an organization is so important because they have such a long and broad view, and a lot of experience from which to draw upon to make their decisions.

Lesson 1-1.2 Business Forecasting and Time Series Data



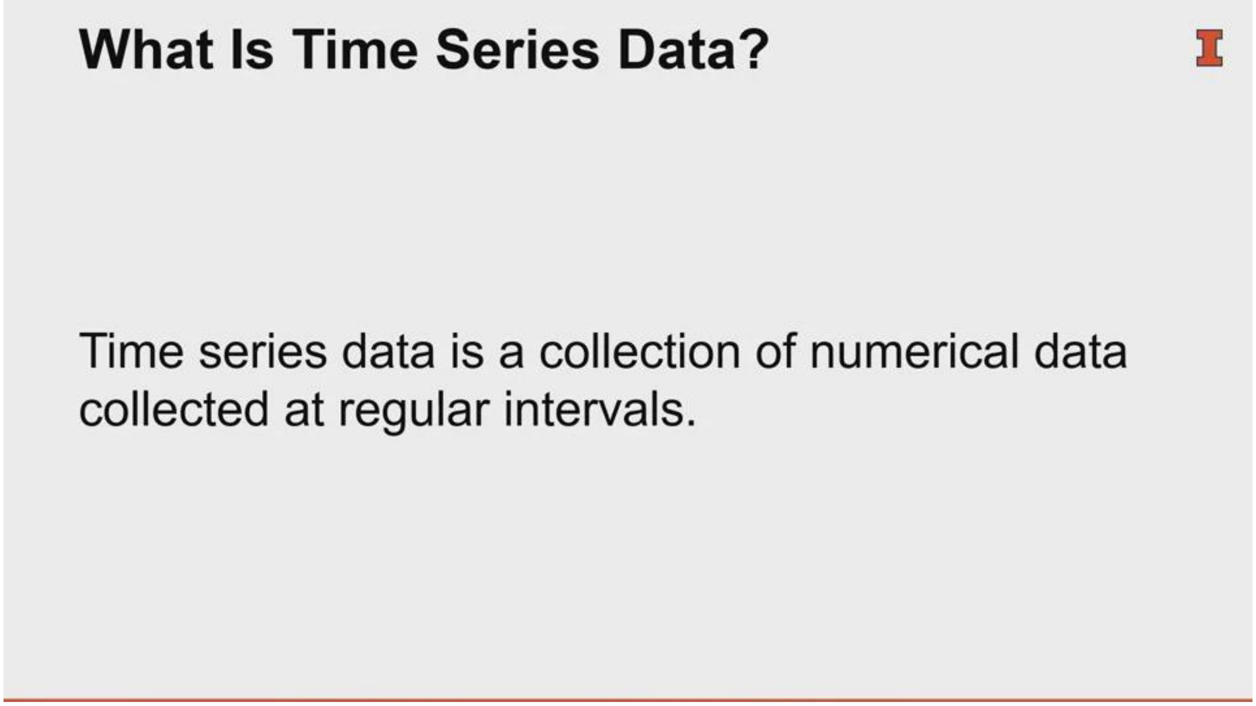
What Is Business Forecasting?

Forecasting is the process of predicting future business events.

Forecasting relies on historical data and forecasting techniques.

Forecasting requires managerial oversight to ensure that the correct forecasting techniques are implemented to prevent costly errors.

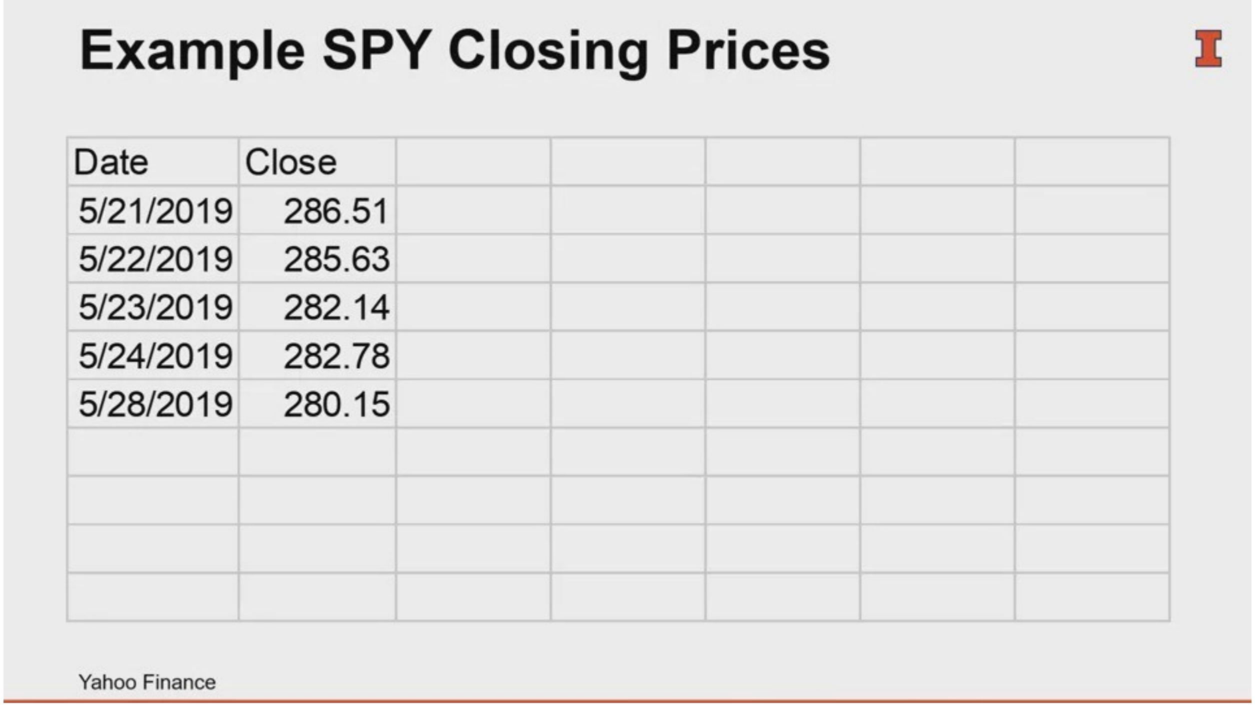
Which brings us to business forecasting and by business forecasting, I'm talking about the process of making forecasts or predicting the future business events, and this forecasting relies on historical data and forecasting techniques. One important thing to note that even though there are these standardized processes or methods to analyze historical data, forecasting does require managerial oversight. So you as a manager will have to ensure that the forecasting techniques are correct, the choice of techniques are correct, and they are implemented appropriately to avoid any constantly errors. So let's begin on some of the empirical notions of business forecasting.
Before we go there, we first need to know what is time series data, time series data.



What Is Time Series Data?

Time series data is a collection of numerical data collected at regular intervals.

Time series data is a collection of numerical data collected at regular intervals, at regular intervals.

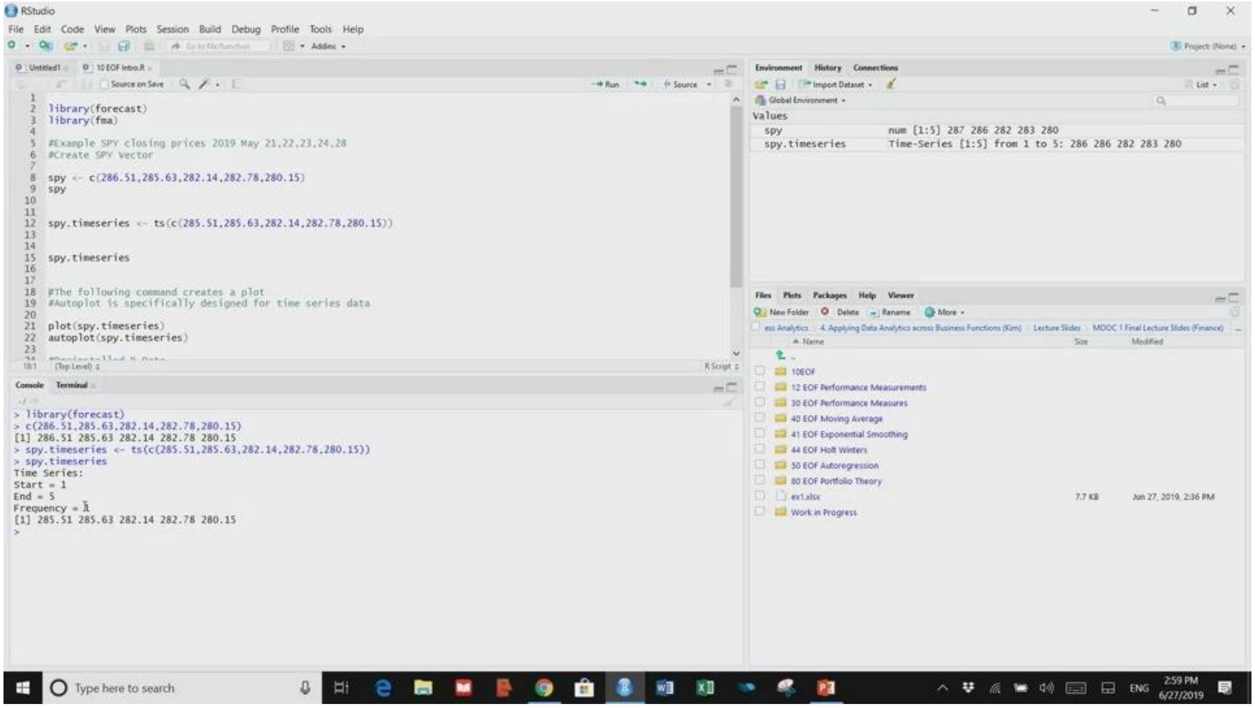


Example SPY Closing Prices

Date	Close					
5/21/2019	286.51					
5/22/2019	285.63					
5/23/2019	282.14					
5/24/2019	282.78					
5/28/2019	280.15					

Yahoo Finance

We have here an example of the closing prices from the SPY from May 2019 the 21st through the 28th. As you can see this is an example of five data points. They're closing prices. It's time- series data. So we have data from the 21st, the 22nd, 23rd, the 24th, and the 28th. They are at regular intervals, they're daily prices. But note here that from the 24th to the 28th is a long weekend here in the United States so there's a three-day gap. But for all intents and purposes they're regularly sequential pieces of data. We don't want to use data that has big gaps like a month missing of daily data or gaps in yearly data and things like that because that would skew our results. So the important thing here is that the data is occurring at regular time intervals.



```
library(forecast)
library(fma)

#Example SPY closing prices 2019 May 21,22,23,24,28
#Create SPY Vector

spy <- c(286.51,285.63,282.14,282.78,280.15)
spy


spy.timeseries <- ts(c(285.51,285.63,282.14,282.78,280.15))


spy.timeseries


#The following command creates a plot
#Autoplot is specifically designed for time series data

plot(spy.timeseries)
autoplot(spy.timeseries)
```

```
> library(forecast)
> c(286.51,285.63,282.14,282.78,280.15)
[1] 286.51 285.63 282.14 282.78 280.15
> spy.timeseries <- ts(c(285.51,285.63,282.14,282.78,280.15))
> spy.timeseries
Time Series:
Start = 1
End = 5
Frequency = 1
[1] 285.51 285.63 282.14 282.78 280.15
>
```

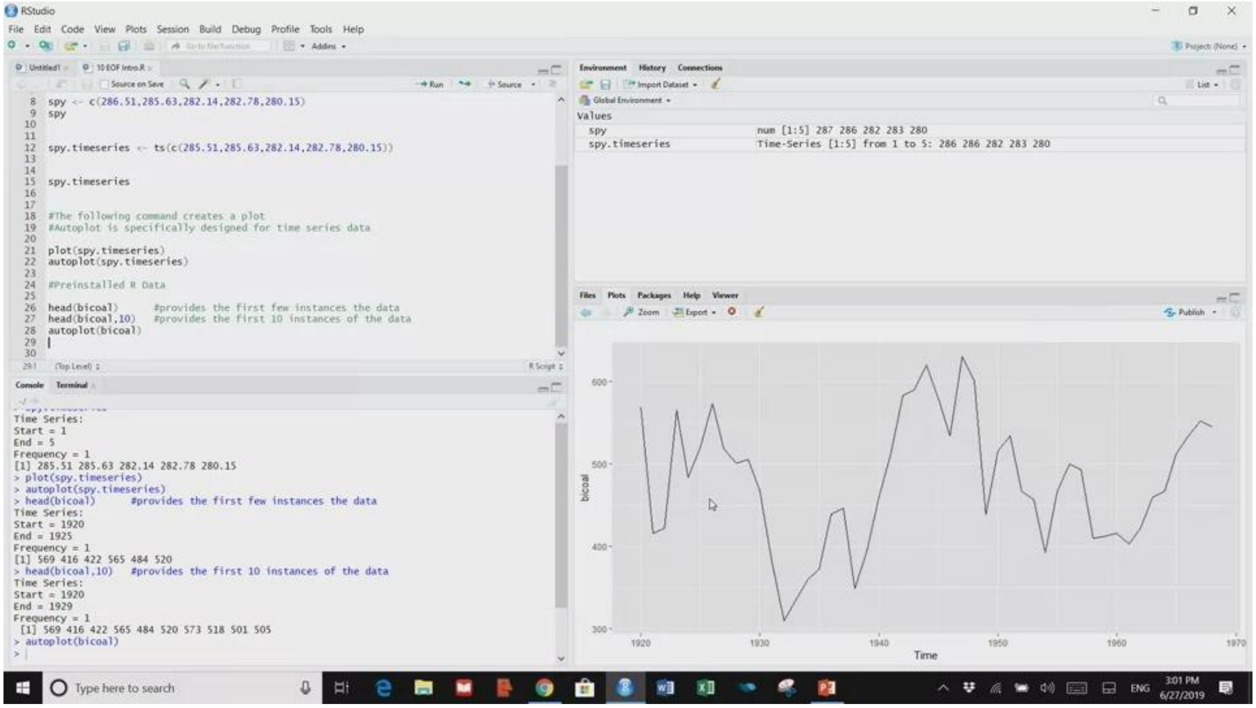
```
Values
  spy               num [1:5] 287 286 282 283 280
  spy.timeseries    Time-Series [1:5] from 1 to 5: 286 286 282 283 280
```

ess Analytics > 4. Applying Data Analytics across Business Functions (Kim) > Lecture Slides > MOOC 1 Final Lecture Slides (Finance)

Name	Size	Modified
..		
10EOF		
12 EOF Performance Measurements		
30 EOF Performance Measures		
40 EOF Moving Average		
41 EOF Exponential Smoothing		
44 EOF Holt Winters		
50 EOF Autoregression		
80 EOF Portfolio Theory		
ex1.xlsx	7.7 KB	Jun 27, 2019, 2:36 PM
Work in Progress		

So here is the example in R. The first two lines of code here are to load the two libraries. One is the forecasting libraries that has the commands for a lot of time series analysis, and the FMA data which has a lot of datasets for us to play with. Then from that table, let's run this. So that'll load the Libraries. I had these prices that I showed you in the slides of the five closing prices and I put them here in a column vector and I put them into a variable named spy. So I create the column vector, I hit Enter, there it is. As you can see down below, there are the five prices. We have here a small little R program to demonstrate what is a time series and how to create one in the R programming environment. The first two lines here are our libraries. The first one loads commands associated with time-series data, and the second one has a lot of data examples for us to play with. So here is the vector of closing prices. As you can see, I typed them into a column vector and I'm going to put that into the variable spy. I run the code and there it is.

But in time-series data, it's sometimes nice to take a column vector and put it into a data structure called a time series. The command for that is simply ts and then you put the data inside the parentheses. We're going to name it spy time series, and there it is. Now we can look at it. The difference between the column vector appear versus the time series data. Is that it gives you a start and end and the frequency which is a one. A one usually means regular intervals periodical versus four or four which means quarterly, etc.



```
 8  spy <- c(286.51,285.63,282.14,282.78,280.15)
 9  spy
10
11
12  spy.timeseries <- ts(c(285.51,285.63,282.14,282.78,280.15))
13
14
15  spy.timeseries
16
17
18  #The following command creates a plot
19  #Autoplot is specifically designed for time series data
20
21  plot(spy.timeseries)
22  autoplot(spy.timeseries)
23
24  #Preinstalled R Data
25
26  head(bicoal)      #provides the first few instances the data
27  head(bicoal,10)   #provides the first 10 instances of the data
28  autoplot(bicoal)
29  |
30
```

```
Values
  spy                    num [1:5] 287 286 282 283 280
  spy.timeseries         Time-Series [1:5] from 1 to 5: 286 286 282 283 280
```

```
Time Series:
Start = 1
End = 5
Frequency = 1
[1] 285.51 285.63 282.14 282.78 280.15
> plot(spy.timeseries)
> autoplot(spy.timeseries)
> head(bicoal)      #provides the first few instances the data
Time Series:
Start = 1920
End = 1925
Frequency = 1
[1] 569 416 422 565 484 520
> head(bicoal,10)   #provides the first 10 instances of the data
Time Series:
Start = 1920
End = 1929
Frequency = 1
 [1] 569 416 422 565 484 520 573 518 501 505
> autoplot(bicoal)
> 
```

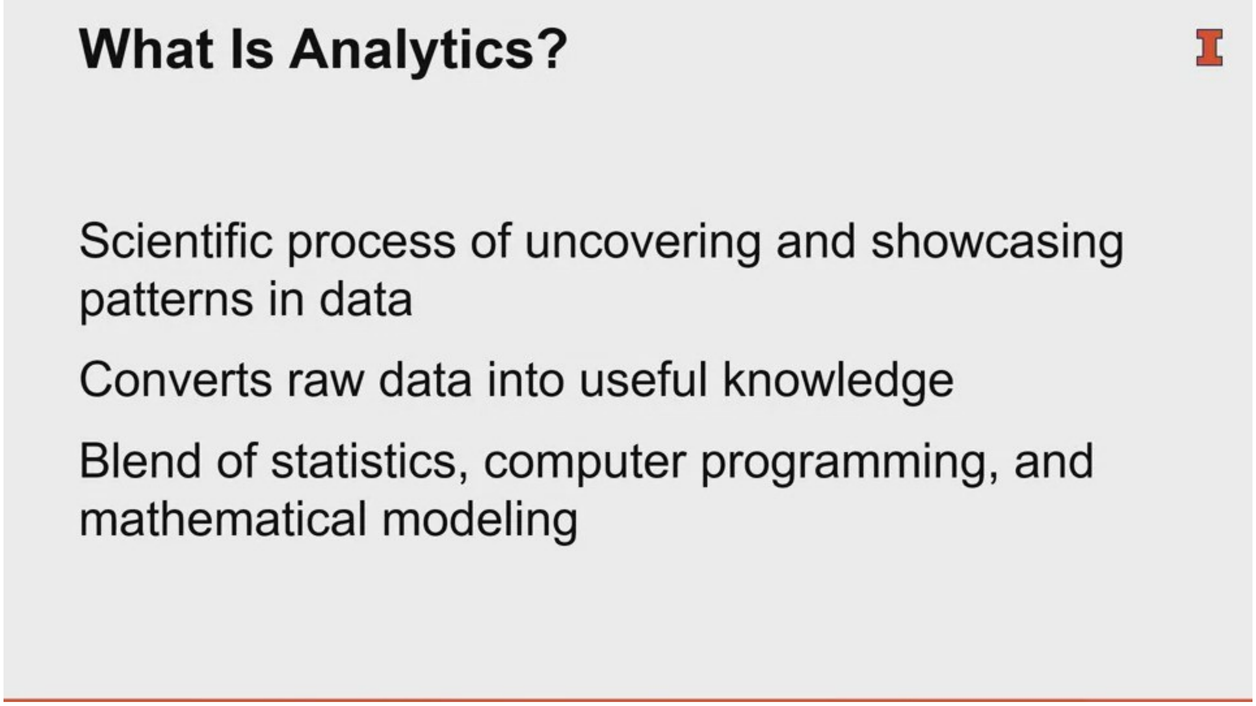
So let's make a plot. We can use the normal plot command in R and there it is, or we can use this autoplot command, which I'll be using more often, that's works better with time series data. As you can see, it looks pretty similar. There is some aesthetic differences, but some of the functionality as we go forward will be apparent in the autoplot. So there is time series data, and you can see that it's trending downwards. Here's another one for coal that's pre-installed in that library. It is time-series data and here you can see it ranges from 19, 20 to 25. The first five elements and here are the values of coal production. This head command if you have a big dataset will just show you the top five rows. You can specify the number of rows which I did here in line 27. To show the first 10 lines of code, you can change that perimeter as you see fit. Then here is the plot command. So let's take a look at that.

There you go. There's some time series data it occurs at regular intervals and the values are listed on the left hand on the y-axis.

Just from the simple graph, we can notice some things that we might be interested in. We can see a drop-off early in 1932 down here. We can see a peak in 47 up here. It's a little unclear if there is an upward trend or a downward trend over the 50-year period, but just visualizing the data gives us a lot of insight. I've uploaded a link to get the actual raw data if you're interested in looking at the table of data. But that concludes this session about what is time series data.

Lesson 1-2 Introduction to Financial Analytics

Lesson 1-2.1 Introduction to Financial Analytics



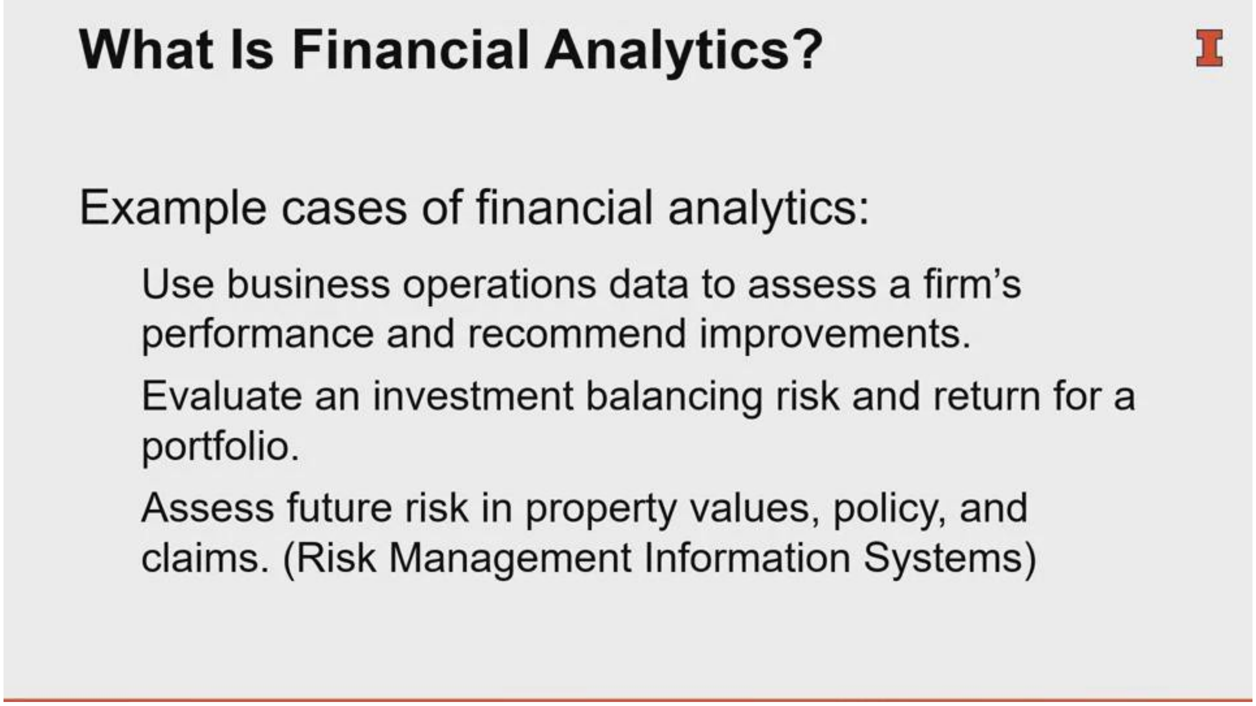
What Is Analytics?

Scientific process of uncovering and showcasing patterns in data

Converts raw data into useful knowledge

Blend of statistics, computer programming, and mathematical modeling

Hello. In this video, I'd like to talk a little bit about what is financial analytics and what we can do with it. So let me just dive right in. In previous classes, I'm sure you've heard about what is analytics. Basically, it's some sort of scientific process for uncovering patterns and showcasing patterns in the data. What the objective of analytics in general is, is to somehow take this raw data and and turn it into useful knowledge, something that we can make decisions on or something to alleviate risk or uncertainty. As I'm sure you're aware of right now, it's a blend of statistics, computer programming, and computer science, and mathematical modeling.

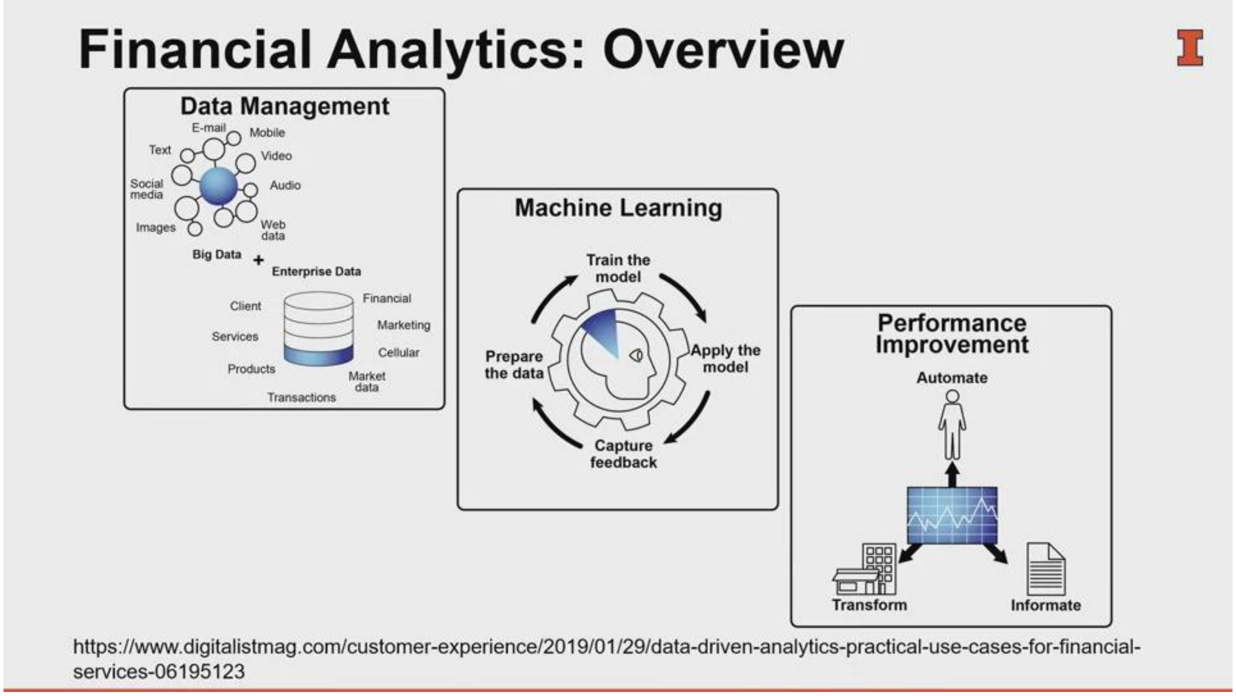


What Is Financial Analytics?

Example cases of financial analytics:

- Use business operations data to assess a firm's performance and recommend improvements.
- Evaluate an investment balancing risk and return for a portfolio.
- Assess future risk in property values, policy, and claims. (Risk Management Information Systems)

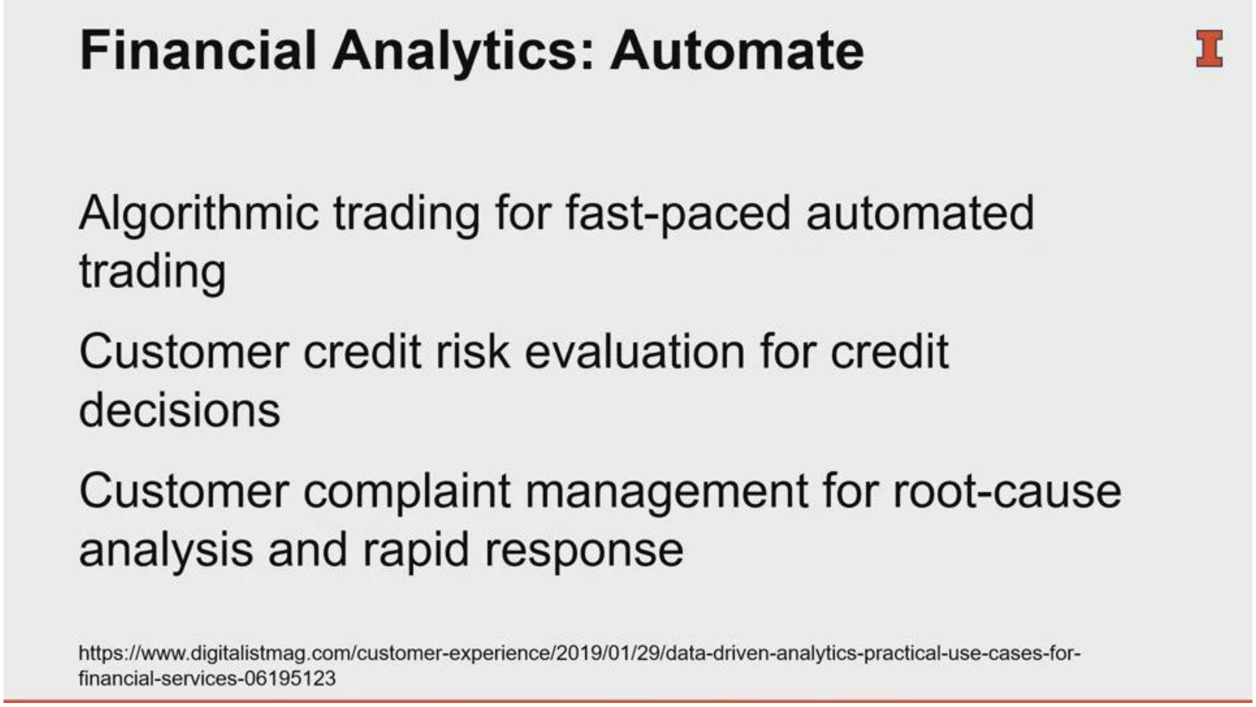
But in the case of financial analytics, what are some examples of uses of analytics in the financial world? One is to use business operations data to assess a firm's performance and recommend improvements. So here, we're mining the data to try to get an understanding of operations to somehow lower costs generally speaking. We might also evaluate investments, trying to balance out risk and reward for portfolio. Another example might be to assess future risk in terms of property value policies and claims. That's often embedded in something that's called a risk management system.



Financial Analytics: Overview

https://www.digitalistmag.com/customer-experience/2019/01/29/data-driven-analytics-practical-use-cases-for-financial-services-06195123

So here's an overview of it. First, we want to manage the data. On the left-hand slide, you can see two major categories of data. At the bottom is enterprise data, you see client data, services, financial marketing data, transactional data. This is traditionally known as enterprise data, and before the advent of the Internet, this is the type of data that people used. It's highly structured, generally contained in some sort of database system, and there was a big thrust. You might have heard of enterprise resource planning systems, ERP systems like SAP, that try to manage this kind of data, generally in-house. With the advent and the rise of social media, the Internet, mobile video, text video, that is augmented enterprise data, and it's now known as big data. Some of the characteristics of big data is the volume of data, there's a vast amount of it, the variety of data. We have email, text, social media. It's also in different formats, not just text-based but audio-based and video-based types of data. So there's a lot of volume and there's a lot of variety. Then the speed or the velocity at which it travels around is also been accelerated with the advent of the internet. So this is the new forms of data, and I think it's the combination of these two terms that really gave popularity or rise to the term analytics in general. But once we have prepared the data and sort of collected it from either internal sources or the external sources, we prepare that data, and we use that to train a model, some sort of mathematical or statistical model. We try to understand what's going on by applying this model hopefully to make better decisions. Once we have acted on that data, we will capture the feedback, what were the results, and then we iterate through this process again, trying to get new data to augment our model, retrain, the model and then keep improving the process. So the question then becomes, why are we doing this? What type of performance improvements are we looking for? There are three forms that I like to talk about. There is automate, informate, and transform, and these are the three types of things you can do with analytics.



Financial Analytics: Automate

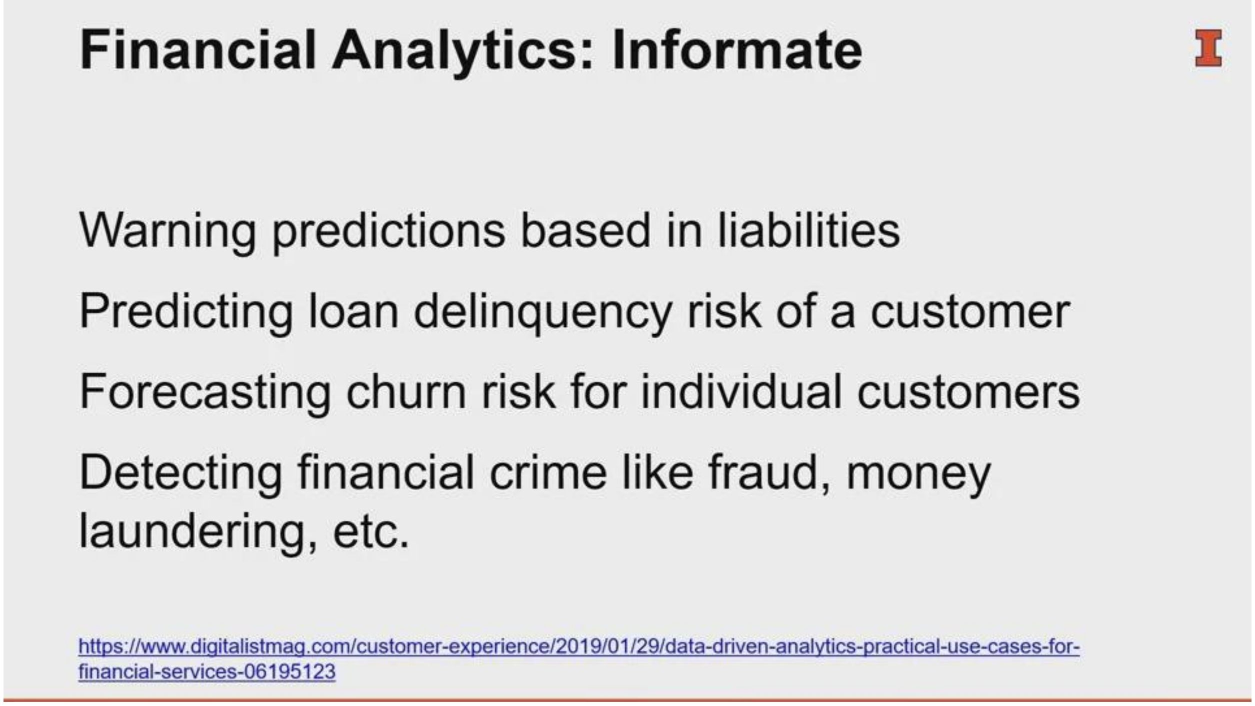
Algorithmic trading for fast-paced automated trading

Customer credit risk evaluation for credit decisions

Customer complaint management for root-cause analysis and rapid response

https://www.digitalistmag.com/customer-experience/2019/01/29/data-driven-analytics-practical-use-cases-for-financial-services-06195123

So what do I mean by automate? Automated processes which automate tasks that are routine, that are easily definable, that are programmable in a computer system. Some examples in the financial world might be algorithmic trading for fast-paced automated trading. So you know there's a movement in the price and your algorithms determines that it's beneficial to buy or sell, to get in there quickly is a competitive advantage. Speed is a competitive advantage. So you can automate these processes. Evaluation of credit risk for customers is another example. Do we give them the $10,000 credit limit for the credit card or not? We use data things like their income, their past credit history, the bank account balance, have they defaulted on loans before. These types of pieces of data are consolidated, and we can automate that decision. More recently, I'm sure you've heard about complaint management for root-cause analysis. Companies are mining they're frequently asked questions to decide what type of problems are occurring in their products or services and how do they manage that process. So that's automate.



Financial Analytics: Informate

Warning predictions based in liabilities

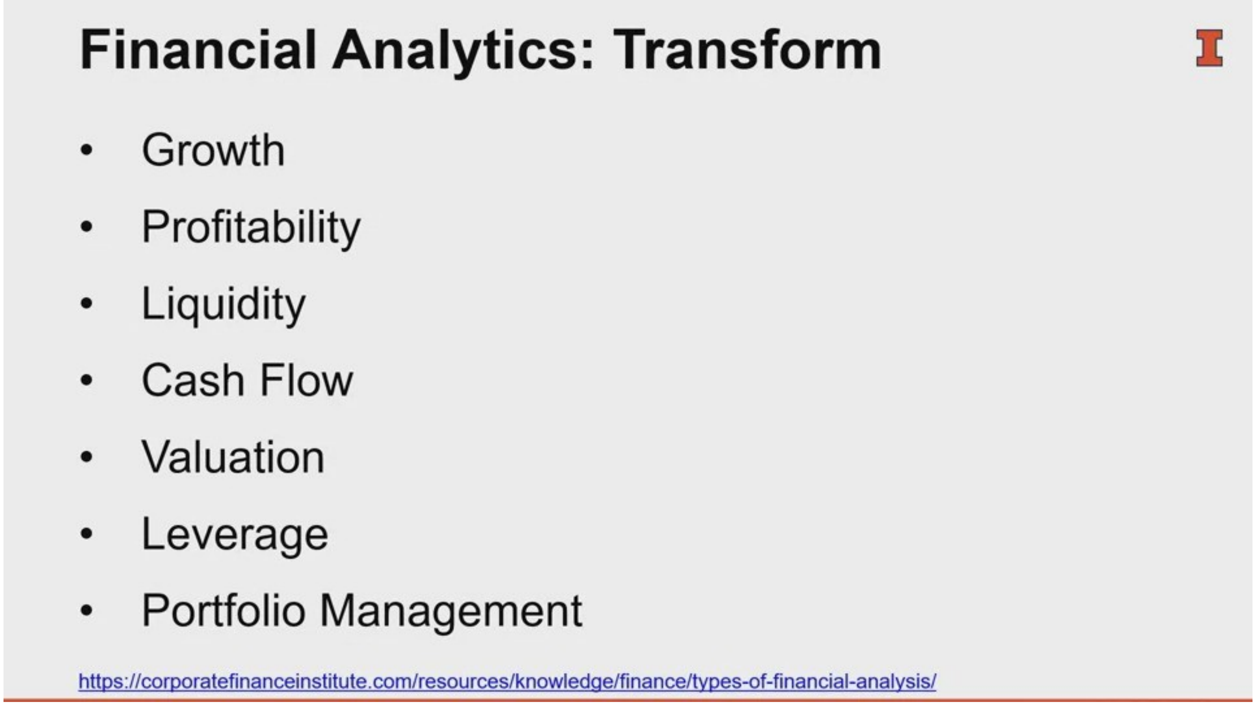
Predicting loan delinquency risk of a customer

Forecasting churn risk for individual customers

Detecting financial crime like fraud, money laundering, etc.

https://www.digitalistmag.com/customer-experience/2019/01/29/data-driven-analytics-practical-use-cases-for-financial-services-06195123

The next category of things we can do with analytics is informate. What I mean by that is somehow collect a large amount of data to make and consolidate it, and summarize it for the decision-makers to use that information along with their knowledge and observations of the world and make an informed decision. So some examples might be, systems that have warning predictions based on liabilities, maybe predicting future loan delinquencies of a customer, trying to understand the nature of the market, and forecasting turn risk for individual customers. Are customers starting to leave your store or running to your competitors? Also used to detect things like financial crime, fraud, money-laundering etc. So all the data's collected and that summary of data is not enough to conclusively determine whether or not something has happened, but an informed decision-maker would be able to use that information to move forward.

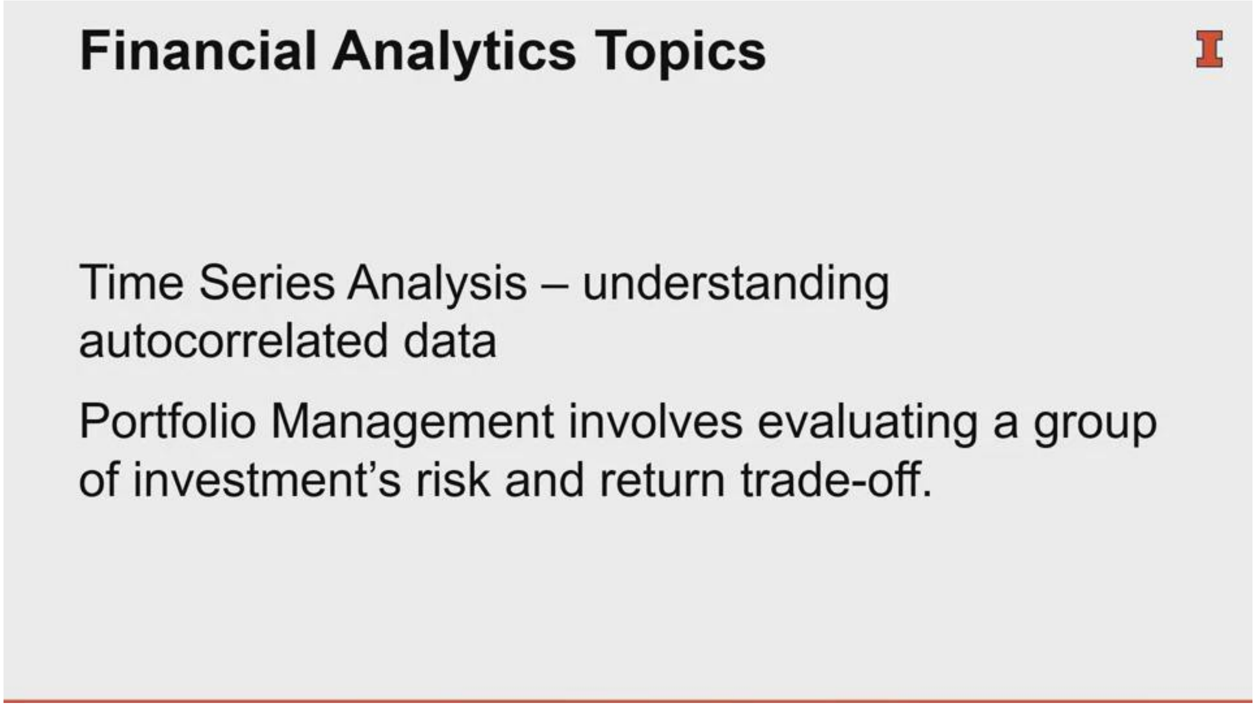


Financial Analytics: Transform

- Growth
- Profitability
- Liquidity
- Cash Flow
- Valuation
- Leverage
- Portfolio Management

https://corporatefinanceinstitute.com/resources/knowledge/finance/types-of-financial-analysis/

Then finally, there's transform. These are things that will just transform the industry completely in terms of growth and profitability, in terms of the financial position, liquidity and cash flow, or valuation. Maybe they're looking for more leverage or somehow maybe managing a portfolio of products, or portfolio of stocks, or a portfolio of companies, and things like that. So to give you a more concrete example, we can think about the transition of Amazon.com and how they went through these three stages of automate, informate, and transform. In the beginning, Amazon was just selling books, and they automated the payment process, they automated the book ordering process. So we didn't need as many telephone operators to answer the phones, to take orders, to write them down, and that was the automate phase of Amazon. Next, we can think about how they use the data in a predictive sense to informate. Once they had this volume of sales data, they were able to predict customer trends and identify things that might sell. They were also able to started to pre-ship items to warehouses closer to where customers might buy things. So for example, seasonal trends on goods, they would move them closer to where the customers want. Then recently, they've been able to take that data and really transform their business in many ways, and I'm sure you can think of many ways, but the one I'm thinking about in particular is that they now have gone into logistical supply chain. They have acquired a fleet of airplanes and delivery vans, and they are now delivering their own products. They know exactly where to go first to save on costs which helps their bottom line or profitability. There are a number of topics in financial analytics, and we can't cover them all in this class. I wish we could.



Financial Analytics Topics

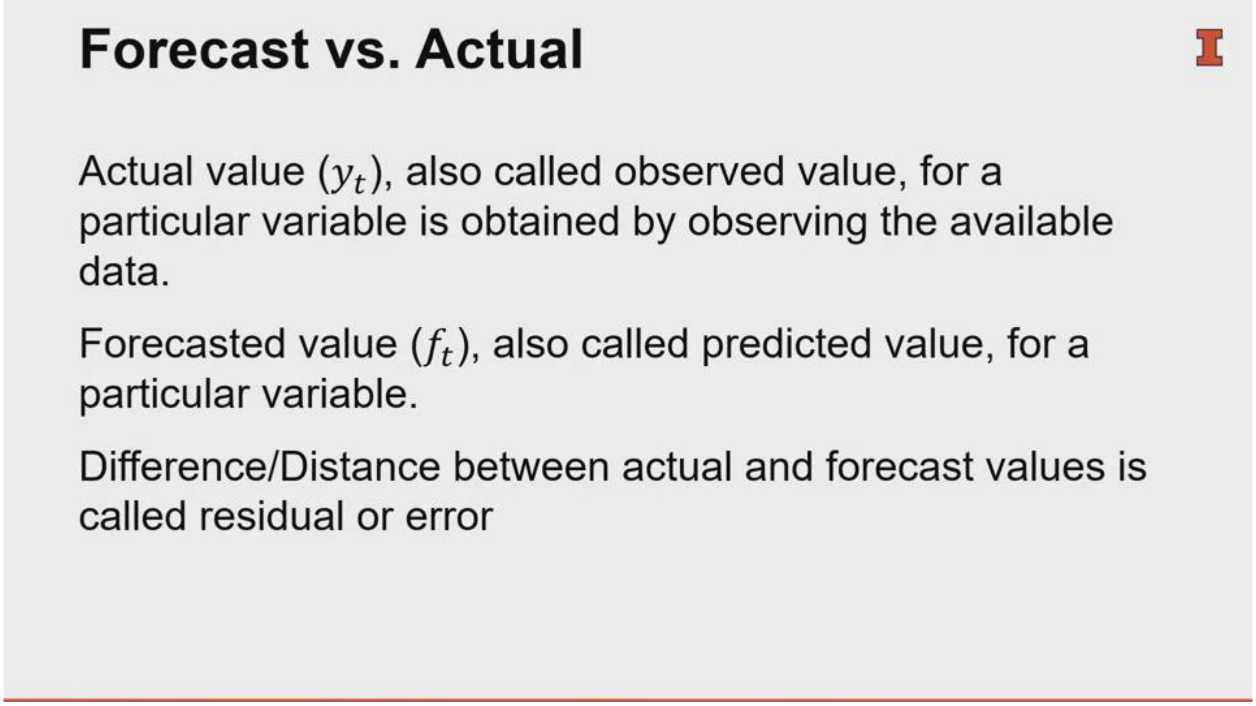
Time Series Analysis – understanding autocorrelated data

Portfolio Management involves evaluating a group of investment's risk and return trade-off.

So in the next four weeks, we're going to cover two broad categories of analytical tools that are often used in financial analytics. First is time series analytics, and that's an understanding of autocorrelated data or data that goes through time. A common example would be the daily prices of stock, daily temperature, noontime temperature is another time series of data, something like that, something that is repeated measures over time. Then second, we're going to look at portfolio management, which involves evaluating a group of stocks for investments based on the risk and return and the trade off between the two. So those are two common tools that we will look at in terms of financial analytics. There are a lot of others, but these are the ones we'll focus on for this class.

Lesson 1-3 Performance Metrics

Lesson 1-3.1 Forecasting Performance Measurements: Distance



Forecast vs. Actual

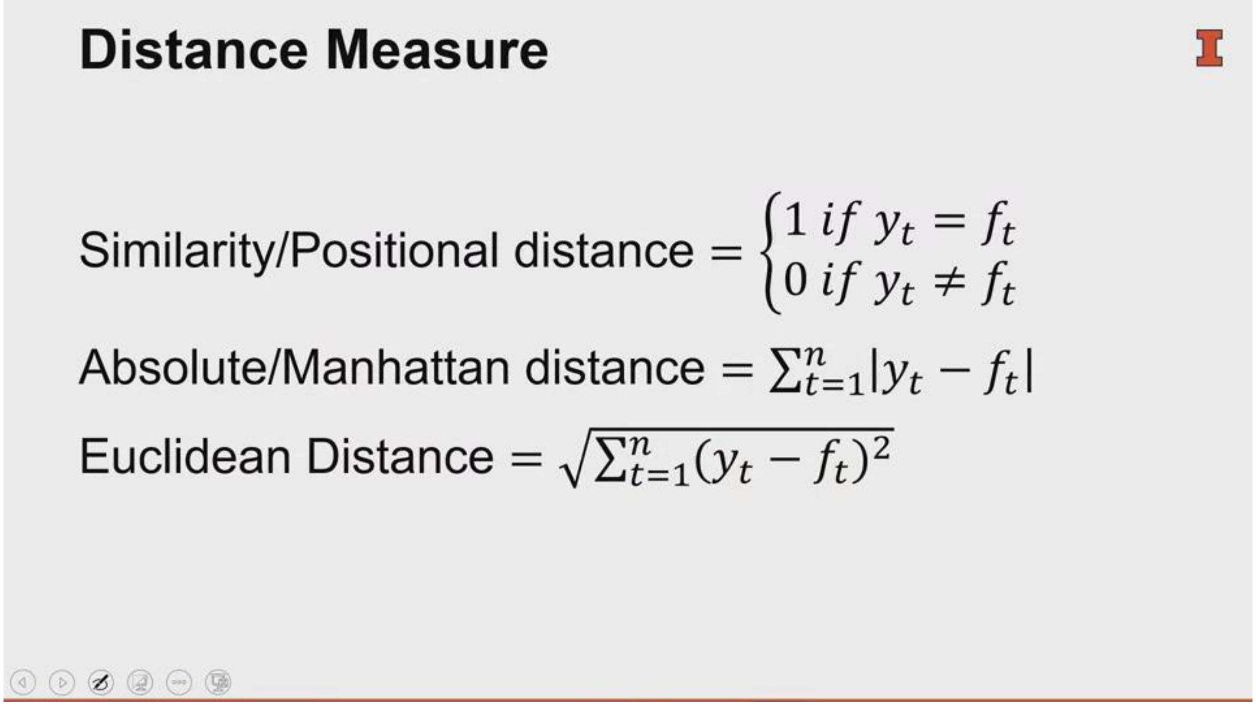
Actual value (y_t), also called observed value, for a particular variable is obtained by observing the available data.

Forecasted value (f_t), also called predicted value, for a particular variable.

Difference/Distance between actual and forecast values is called residual or error

Hello. In this set of slides, I'm going to talk about performance measurements. What are performance measurements? We have seen a couple of ways at least subjective methods of making forecasts. The idea is to figure out a way to benchmark the different methods that we use to make a forecast, which ones are good and which ones are bad? In order to do so, we need to have some performance measurements. Before I talk about the various performance metrics or measurements, I want to talk about a little bit of notation. For actual values, sometimes called observed values for a particular variable, I'm going to denote it as yt. yt is the actual observed values, and we obtain this data through measurements of the real world or the external environment. They could be noon time temperatures. For example, today, the temperature at noon is 75 degrees Fahrenheit, yesterday was 70 degrees Fahrenheit. Those are the actual observed variables. Notice also in here, we have that t, that denotes the time. So if it's a daily time series, would be today, yesterday, two days ago, etc. That's what the t subscript denotes.

Our forecasted value also called a predicted value is our best guess of time period t. We know today's value. So for predicting t plus 1, for example, that would be our forecasted value. The difference between the actual and the forecast is called residual error. That really describes the difference and provides the foundation for our performance metrics. That would be y_t minus ft. One note, if you look at various textbooks, sometimes they will put a hat on f to denote that that's an estimate for that value. Okay. Just now intuitively, I described this y of t minus ft.



Distance Measure

$$\text{Similarity/Positional distance} = \begin{cases} 1 \; if \; y_t = f_t \\ 0 \; if \; y_t \neq f_t \end{cases}$$

$$\text{Absolute/Manhattan distance} = \sum_{t=1}^{n} |y_t - f_t|$$

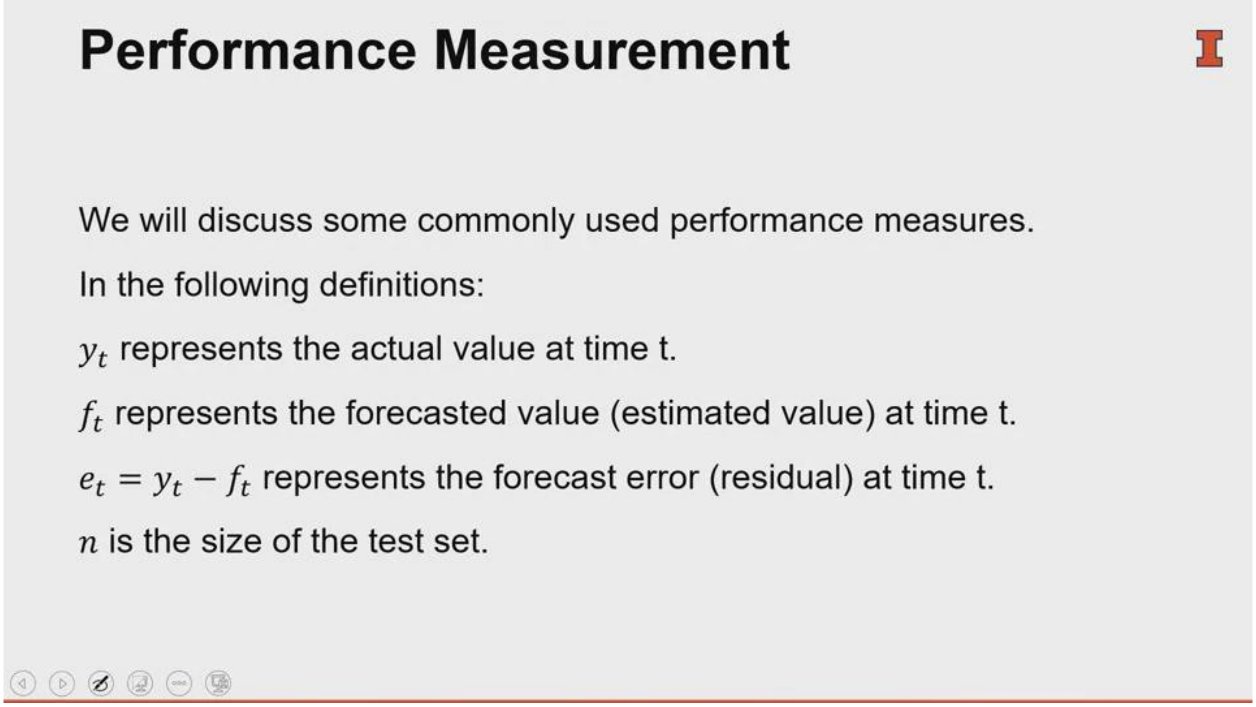
$$\text{Euclidean Distance} = \sqrt{\sum_{t=1}^{n} (y_t - f_t)^2}$$

But there are other ways that we can measure distance. What we're really trying to get is how far apart is my best guess versus the actual measurement? I want to talk about distance measures.

What we're really trying to get at is some notion of how accurate our prediction is from our actual value and our forecasted value. yt minus ft. What's the difference between my actual value and my predicted value? If I predicted today's temperature was going to be 77 degrees Fahrenheit and it was actually 75 degrees Fahrenheit, that's a difference of two degrees. So we can call this distance measures. I want to drill down a little bit about distance measures and be a little more mathematically precise. There are three forms of distance measures that are often used in analytics. The first one is positional distance. Basically, it's either here or not here, they're the same or they're different. If they're the same, there is no distance between them and we'd give it a zero. If they are different, if y of t is different from f of t, then that's 1. This might sound like an odd notion of distance, but if you think about it in your real life, you can think of examples where that might make sense. For example, say you're waiting for a friend at a cafe and they're late and your friend is either here or not here. That's what this idea, positional distance gets at.

The second notion of distance, absolute distance, is something that I think we intuitively grab onto more often than not. It's simply an absolute value. y of t minus f of t, and then it's a distance, so we take away the sign. So if I predict today's noontime temperature to be 75 degrees Fahrenheit, but it was actually 70 degrees Fahrenheit, then I was over by five. But if I predicted 65 degrees Fahrenheit and was actually 70 degrees Fahrenheit, then I'm under by five. We would have to keep track of the sign. So if we're just interested in how far apart they are, we just take the absolute value. Now, the summation here is overtime if I'll make multiple predictions, I want to know the distance how far apart I am for my actual values and my predicted values for all of them. So I just add them up. That's basically what this absolute distance is getting at. Finally, we have this notion of Euclidean distance. It's basically right here is the same, yt minus ft. That's the same thing you see in absolute distance. But instead of using absolute values, we use a squared term, and then we take the square root. This makes sense in multiple dimensions. You might have seen it when you have used the Pythagorean theorem to figure out the length of a hypotenuse of a right-angled triangle. So that's where this notion comes from. In point of fact, this last Euclidean distance is probably the most commonly used in analytics and statistics, partially because of its usefulness, and partially because of its mathematical qualities.

Lesson 1-3.2 Forecasting Performance Measurements: Metrics



Performance Measurement

We will discuss some commonly used performance measures.

In the following definitions:

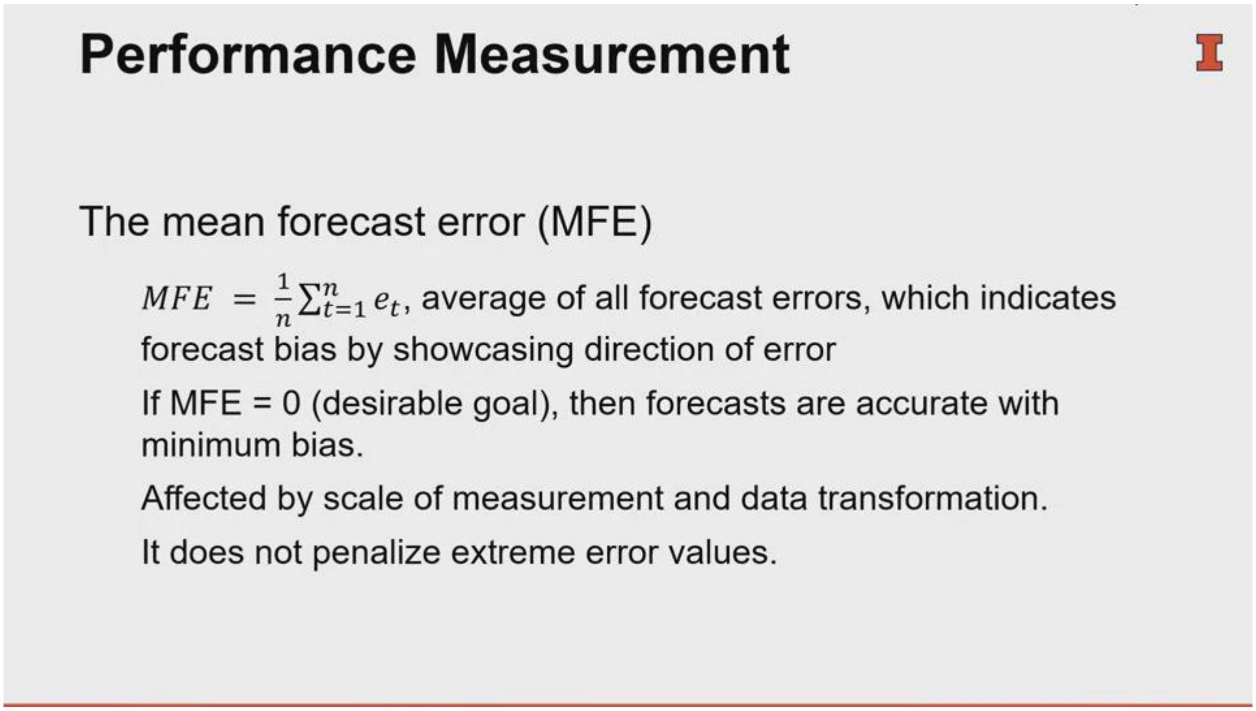
y_t represents the actual value at time t.

f_t represents the forecasted value (estimated value) at time t.

$e_t = y_t - f_t$ represents the forecast error (residual) at time t.

n is the size of the test set.

So let's look at some performance measurements. Again, to summarize the notation, y of t represents the actual value at time t, f of t represents the forecasted value or estimated value at time t, and the error term, the residual e of t is denoted y of t minus f of t. Finally, n is the size of the test set, the size of your time series, how many data points do you have? So if you are looking at a time series of data over 30 days, you'll have 30 days of actual data, and you'll have 30 days of forecast that you made before that event occurred, and then you would have to be able to take the difference.



Performance Measurement

The mean forecast error (MFE)

$MFE = \frac{1}{n}\sum_{t=1}^{n} e_t$, average of all forecast errors, which indicates forecast bias by showcasing direction of error

If MFE = 0 (desirable goal), then forecasts are accurate with minimum bias.

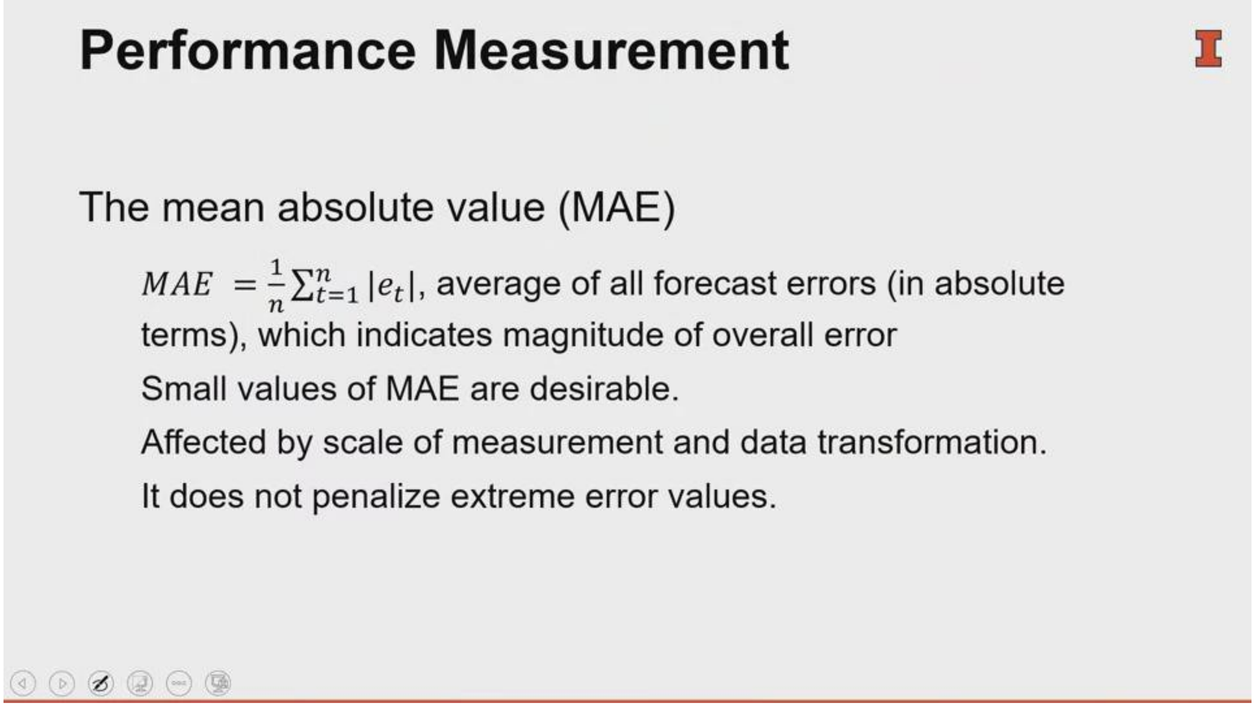
Affected by scale of measurement and data transformation.

It does not penalize extreme error values.

The first performance measurement that I would like to talk about is called the mean forecast error or MFP for short. The formula is noted here. Let's drill down a little bit on that formula. E of t is equal to, as noted on the previous slide, y of t minus f of t. Then we have this summation term that adds them all up. Then we divide by n. If you look at this carefully, it's very much like an average. In fact, it is an average. It's an average of your error terms. One thing I want to point out is, how do you read these formulas and translate that into a spreadsheet? So let's say you have a spreadsheet and you have your column heading, and then you have some data. When it says, you see that giant Sigma, that's the summation sign. But I want you to think is that e of t is just a column of data, and you're just going to add them all up. So you have a column of data, and you're just going to add them all up. That's what this portion here represents right here. Okay. That's a column of data and added up. Nine times out of 10 when looking at these formula, you're going to add up the whole column.

2:19

N is the number of observations. So t goes from one to n, so we're going to add up the column, and then we're going to divide by n. So that is basically your average of your error term. Another way to think about this since we have this y_t and f of t column, is that you have y of t, and you might have some observations. You have f of t in your spreadsheet, and you have sums observations, and in this column, you have e of t where you take the corresponding values y of t minus the corresponding values of f of t, and then you take the difference, and then you just add them up and divide by n. There are some properties regarding the mean forecast error. Ideally, zero is the best, and we can see that here. If y of t and f of t are identical, meaning your prediction hit the mark and it's exactly the same as the actual value, that term will be zero. The more zeros you have, the close you are to zero. So that's the desirable goal. It is affected by the scale of the measurement and data transformation. So the actual nominal value, the actual numerical value will be affected by things like using Fahrenheit as a scale, or Celsius as a scale. The numbers mean something different. So that's something to keep in mind. The other thing it does not do is penalize for extreme errors. So if you make a forecast that was way off, it doesn't really penalize for that.



Performance Measurement

The mean absolute value (MAE)

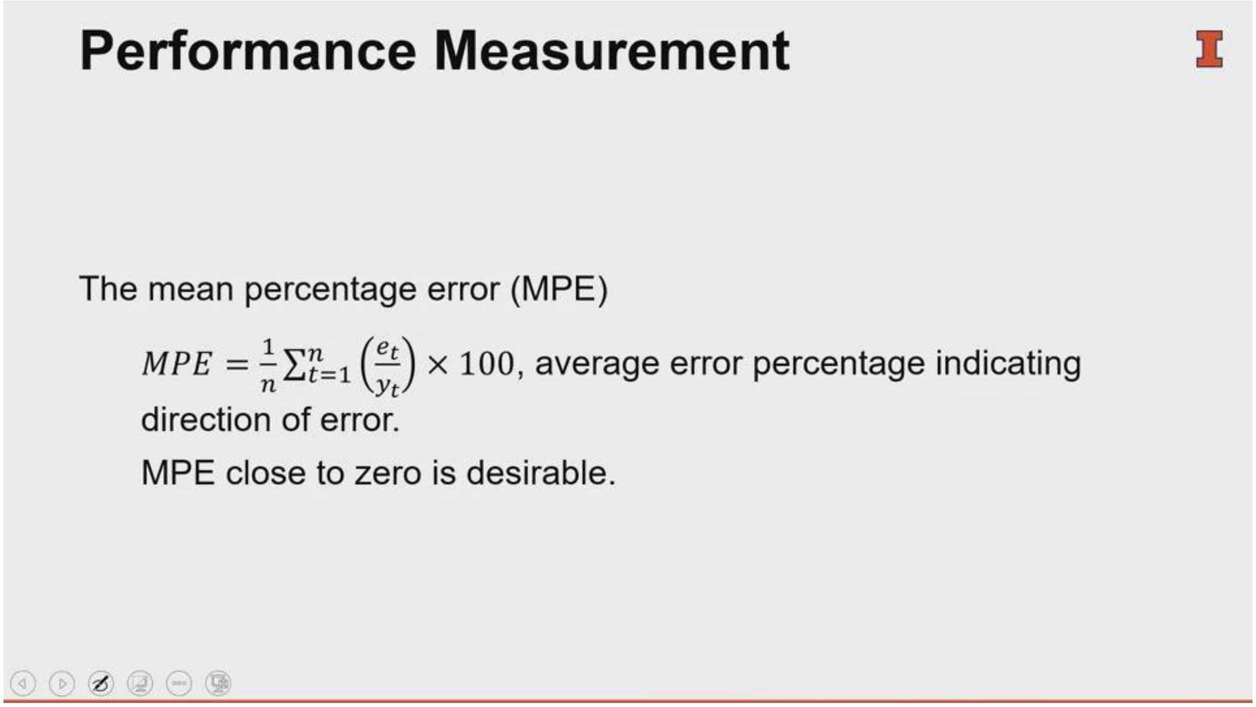
$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|$, average of all forecast errors (in absolute terms), which indicates magnitude of overall error

Small values of MAE are desirable.

Affected by scale of measurement and data transformation.

It does not penalize extreme error values.

The next measurement I want to talk about is the mean absolute value. Here again is that, e of t term, and like before, it's the y of t minus f of t is equal to e of t. But instead just adding them up, it takes the absolute value and divides by n. So it strips the sign. So it's just looking at how far away the prediction is. It doesn't matter if it's above or below. Okay. As before, small values of the mean absolute value are desirable, and again, it is affected by measurement or data transformations, and again, it does not penalize for extreme values.



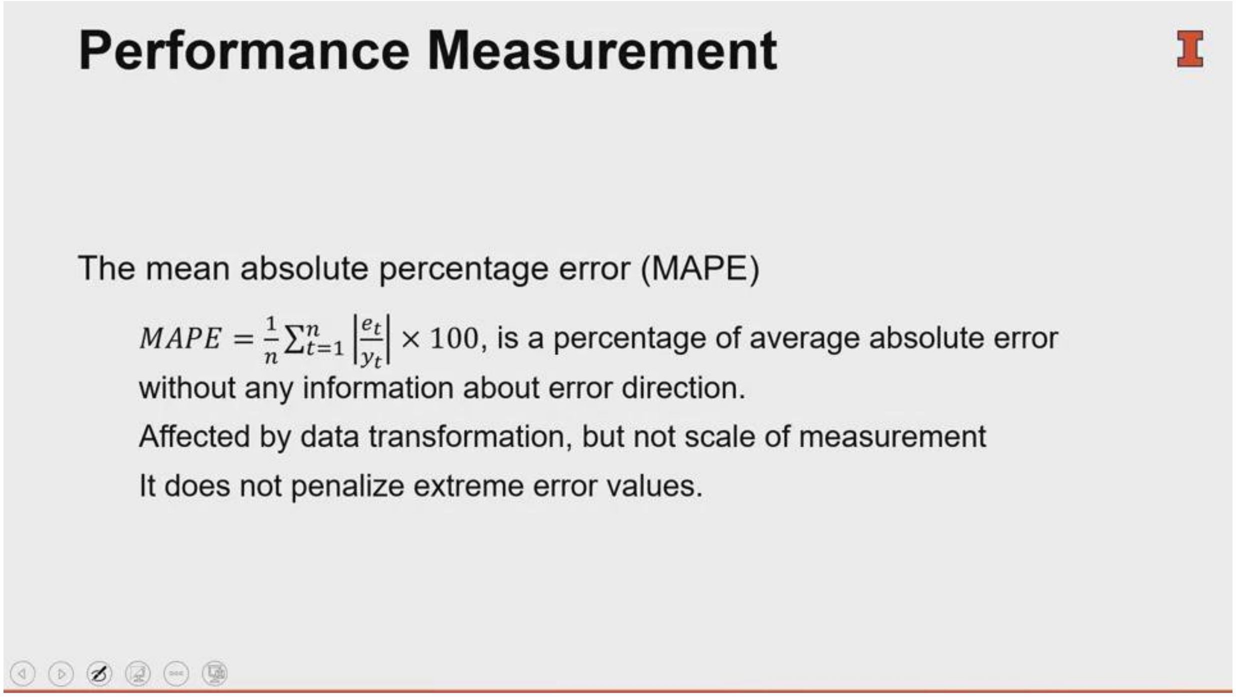
Performance Measurement

The mean percentage error (MPE)

$MPE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{e_t}{y_t}\right) \times 100$, average error percentage indicating direction of error.

MPE close to zero is desirable.

Another one is to take percentages. So instead of taking the absolute difference, we're going to look at percentages. To get the percentage, we take the difference between our forecast value and the observed value, and we divide by the observed value. That's our forecast. We do that for every time period, and then we take the average multiply it by a 100 to get a percentage. Again, like the other ones, being close to zero is desirable. That's again because of this e term, which is essentially y_t minus f of t. We want those to be on top of each other.



Performance Measurement

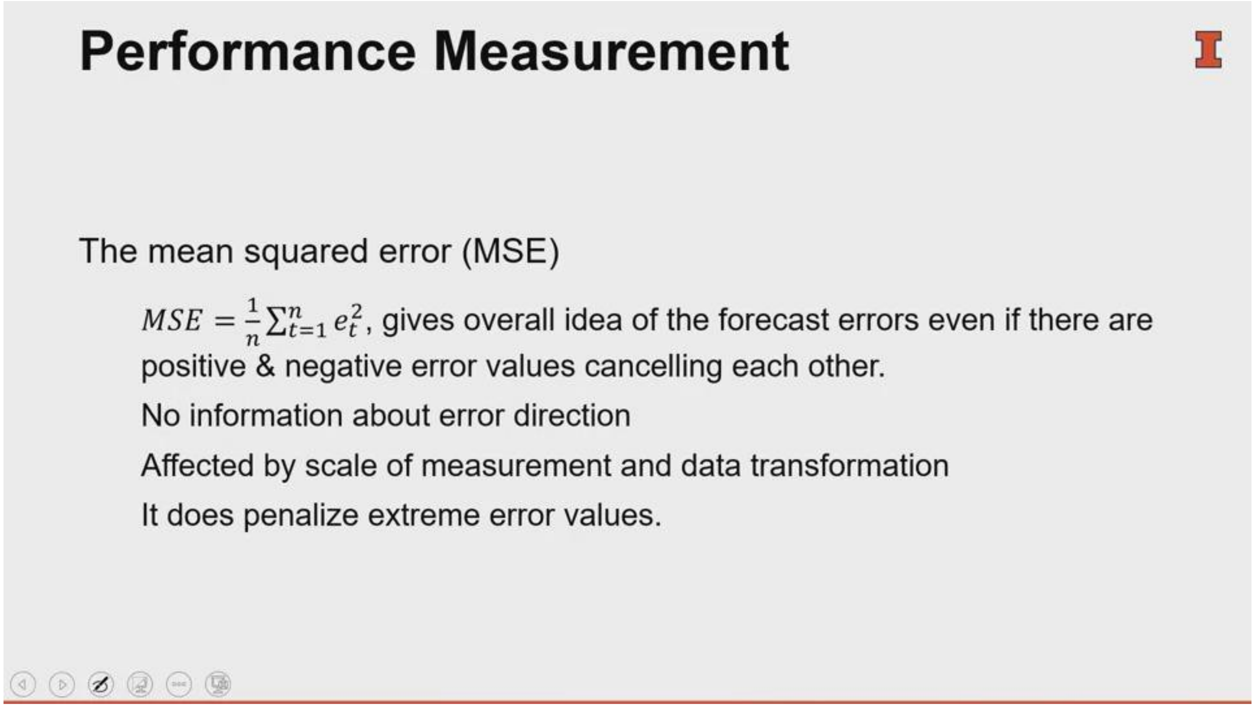
The mean absolute percentage error (MAPE)

$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \times 100$, is a percentage of average absolute error without any information about error direction.

Affected by data transformation, but not scale of measurement

It does not penalize extreme error values.

We can take absolute percentage value, is very similar to the mean percentage error term, but here we just take the absolute value. We're stripping away the sign. Again, being close to zero is better than being far apart.



Performance Measurement

The mean squared error (MSE)

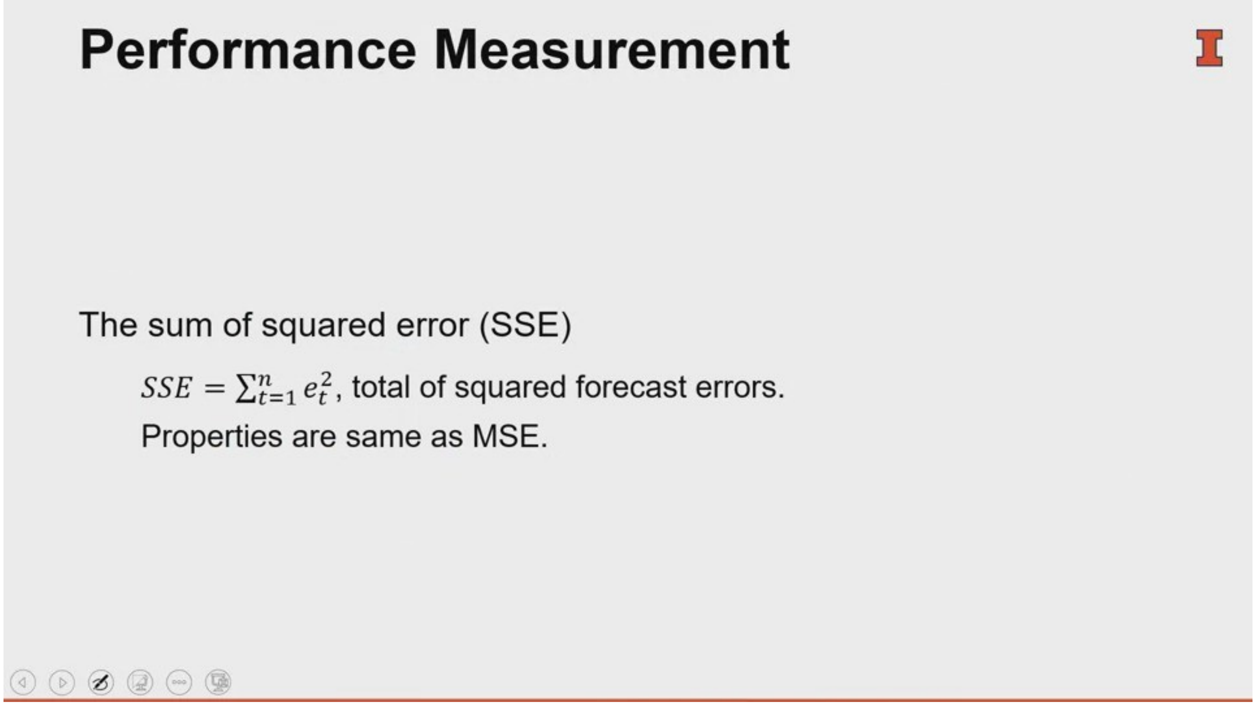
$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$, gives overall idea of the forecast errors even if there are positive & negative error values cancelling each other.

No information about error direction

Affected by scale of measurement and data transformation

It does penalize extreme error values.

Those last terms just looked at the absolute distances. Here, we're looking at squared error, the Euclidean distance, and so here we have this e sub t squared, which is y_t minus my forecast value squared. So again, you have some data, we have my forecast, and we have some data in a spreadsheet, and then a third column we take y_t minus f of t. So we have this stuff, then we square that term, and we add them up, and then we take the average of those. Note, there is no direction about the error direction because by squaring the term, all the values are positive. It is affected by the scale of measurement and the data transformation. But in this case, it does penalize for extreme error values.



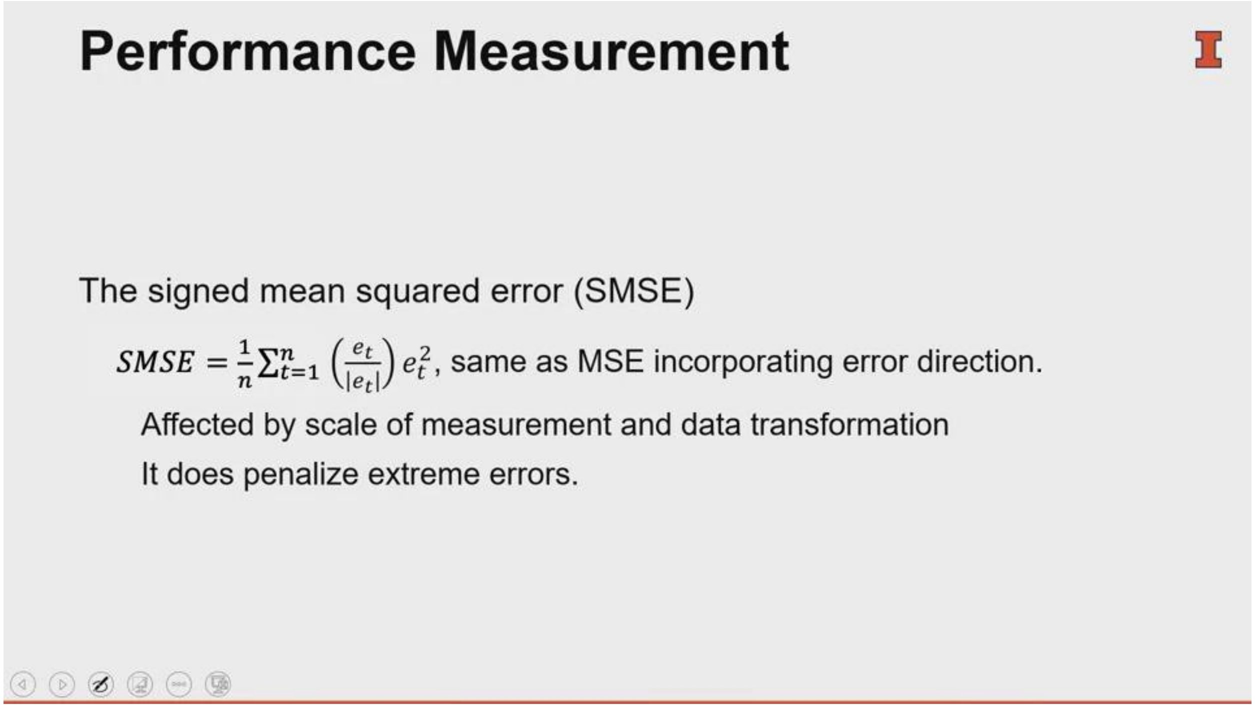
Performance Measurement

The sum of squared error (SSE)

$SSE = \sum_{t=1}^{n} e_t^2$, total of squared forecast errors.

Properties are same as MSE.

Related to mean squared error, is sum of squared error, sometimes denoted SSE. Here we have simply the sum of all those numbers. It has the same measurement characteristics or properties as the mean squared error, and x is the signed mean squared error.



Performance Measurement

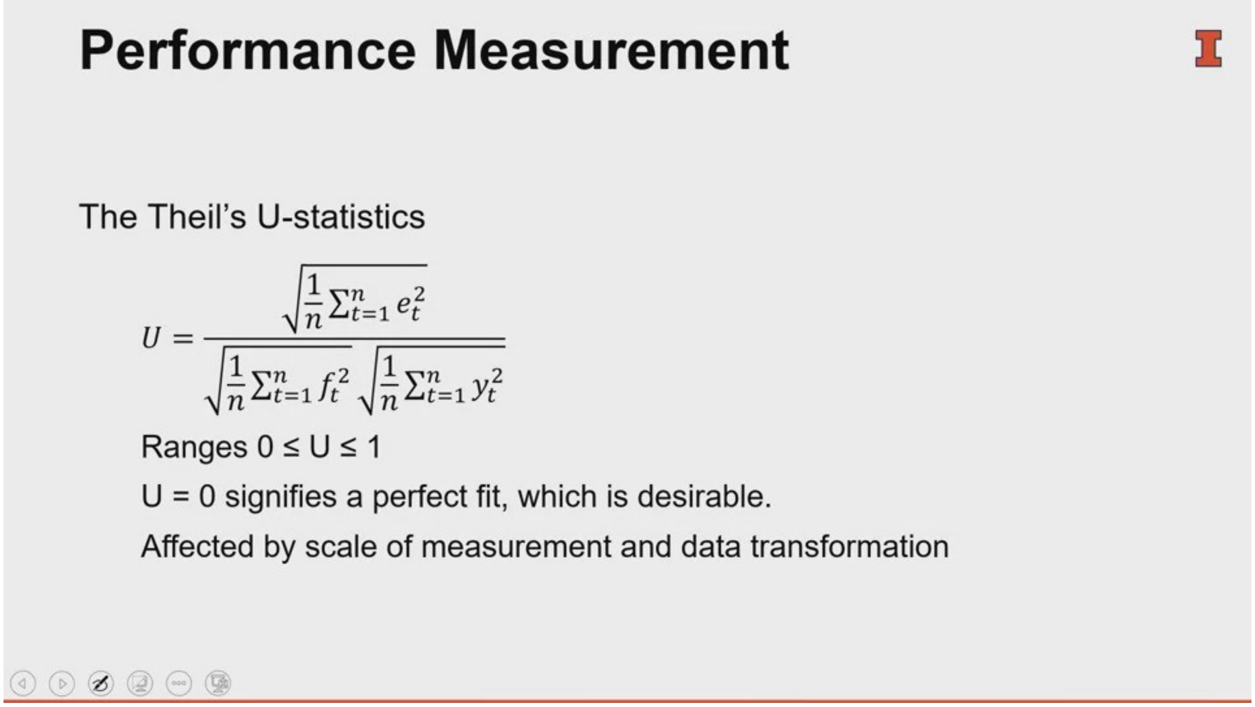
The signed mean squared error (SMSE)

$SMSE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{e_t}{|e_t|}\right)e_t^2$, same as MSE incorporating error direction.

Affected by scale of measurement and data transformation

It does penalize extreme errors.

So here we have the same term, the e_t squared, but here is where we get the sine part, the signed mean squared error. So if we look at this term e_t over the absolute value of e_t, this numerator will either be plus or minus, and this denominator will always be positive. So that's how the sign of the term gets carried in there. So it adds or subtracts as we go through these e_t squared components. That's how it incorporates the direction of error, whether it's too high or too low. Unlike before, the squared error terms does penalize for extreme errors.



Performance Measurement

The Theil's U-statistics

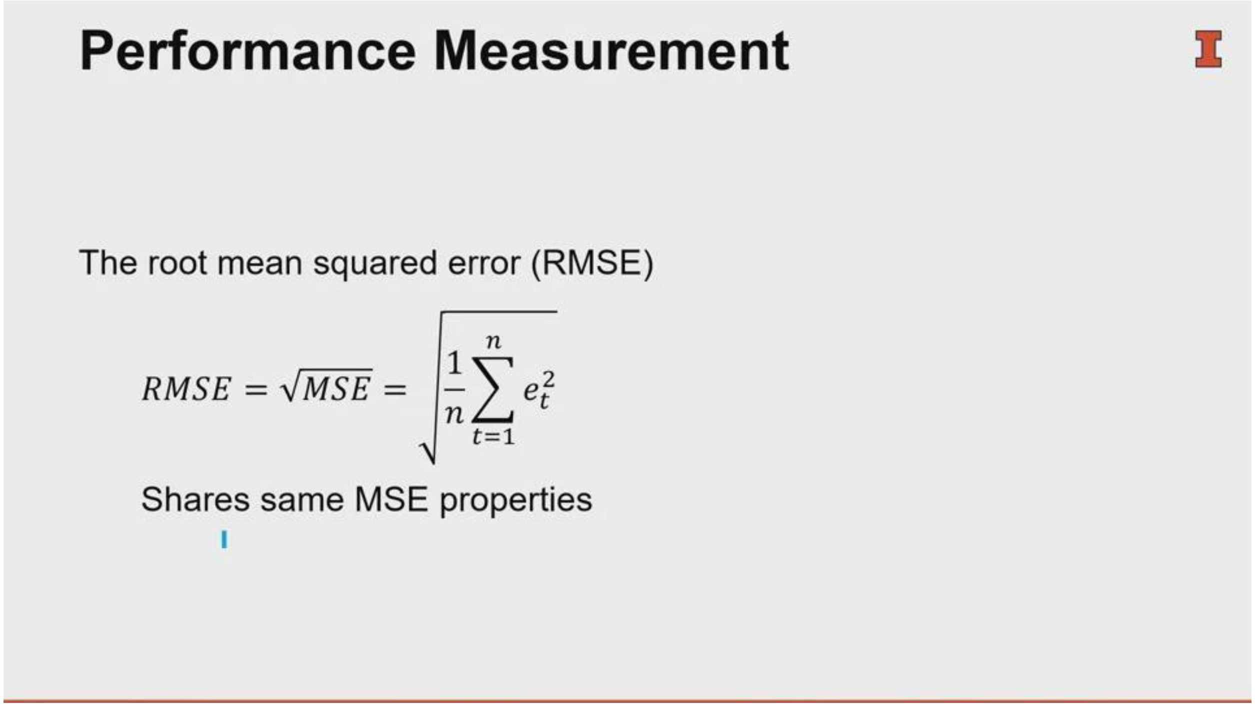
$$U = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n} f_t^2}\ \sqrt{\frac{1}{n}\sum_{t=1}^{n} y_t^2}}$$

Ranges 0 ≤ U ≤ 1

U = 0 signifies a perfect fit, which is desirable.

Affected by scale of measurement and data transformation

Here's another one. It's called Theil's U-statistic. It's a complicated formula, but we can break it down. Here's the mean squared error, and then it takes a similar calculation with the forecasted value and the actual values. So the errors are divided by these two terms. The U-statistic ranges from zero to one, where again U equaling zero represents a perfect fit. That's right here. In this part of the equation, again y_t minus f of t is e_t, and then if we square that t squared term and we add them up and take the average. Again, if this is zero, then the whole numerator is zero, and so U become zero. Otherwise at the other end of the spectrum, it's a one. So we want small values of this final one.



Performance Measurement

The root mean squared error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

Shares same MSE properties

Finally, I'm going to talk about the root mean squared error, which is essentially the mean squared error, you can see it, and you take the square root. It shares the same properties of the mean squared error, and this is the probably the one that we will use most often in this course. In general, this is a very popular performance metric to use. So that really concludes our section on performance measurements. We've talked about a number of them. We talked about distance. In the end, I want you to focus on this one in an applied science. As you gain more experience, you'll see some other ones that are also quite useful. That wraps up performance measurements.