Advanced Data & Network Mining

Specialized Applications
Time Series Forecasting

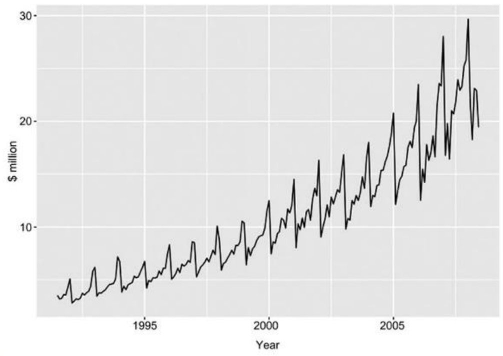
2023-24

Data Mining
Time Series Analysis and Forecasting

Kotu and Deshpande (2019) describe time series as series of observations (datapoints) listed in constant successive time intervals.

Time series **analysis** - process of extracting meaningful non-trivial information and patterns from time series.

Time series **forecasting** - process of predicting the future value of time series based on past observations and other inputs. One of the oldest known predictive analytics techniques, widely used in every organizational setting with deep statistical foundations.



Time series of monthly antidiabetic drug sales.[1]

(Left) Time Series of Antidiabetic Drug Sales (Kotu and Deshpande, 2019, p.396)

Data Mining
Time Series Analysis and Forecasting

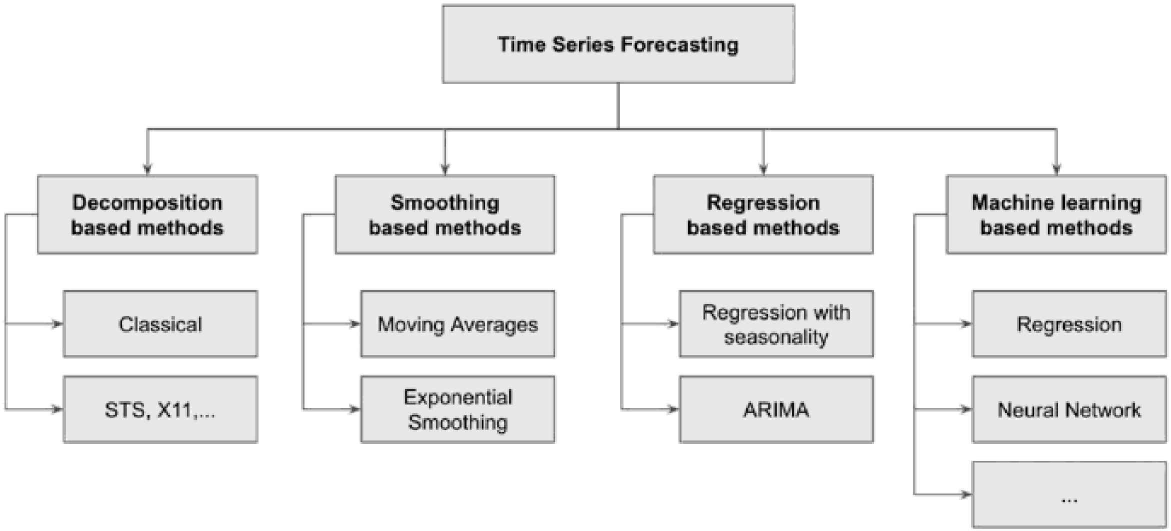
The supervised model building to date has been was about collecting data from several different attributes of a **system** and using these to fit a **function** to predict the desired quantity or target variable. **Predictors or independent variables are used to predict the target variable.**

- housing market - price of a house, square footage, number of bedrooms, number of floors, and age used to predict median house value.
- pricing - data from several different influencing commodity prices used to model the cost of a product.

The objective in time series forecasting is different. **Historical information about a particular quantity is used to make forecasts about value of the same quantity in the future.**

Univariate time series forecasting does not need the independent or predictor variables used for multivariate time series.

Data Mining
Time Series Forecasting Taxonomy



Taxonomy of Time Series Forecasting
(Kotu and Deshpande, 2019, p.398)

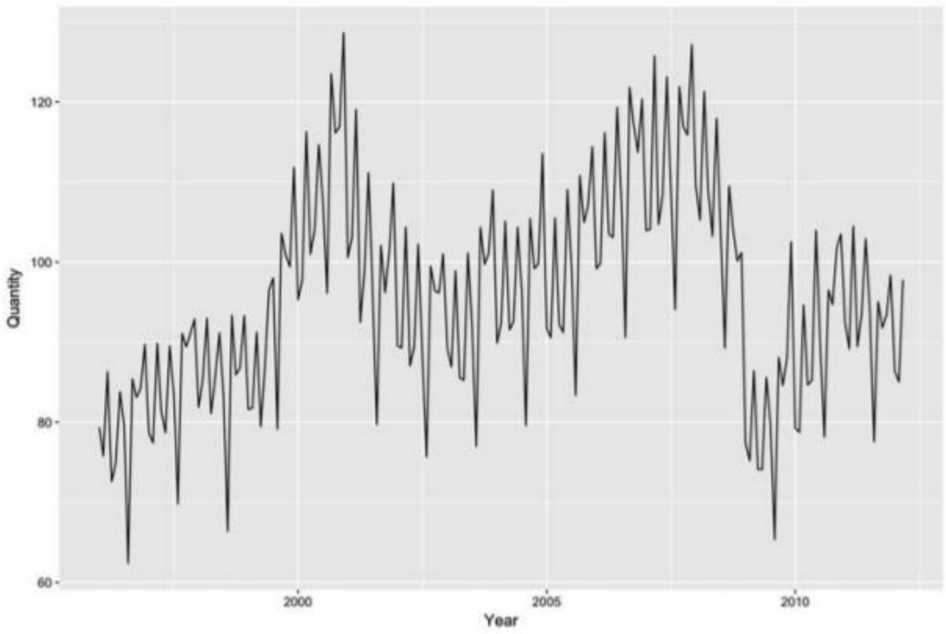
Time Series Forecasting : Case Study
Forecasting demand of a product

A very common application of time series is in forecasting demand for a product. A manufacturing company makes anticorrosion wax tapes for use in gas and oil pipelines. The company makes more than a dozen varieties of wax tape products using a handful of assembly lines. The demand for these products varies depending upon several factors. For example, routine pipeline maintenance is typically done during warm weather seasons. So there could be a seasonal spike in the demand. Also over the last several years, growth in emerging economies has meant that the demand for their products also has been growing. Finally, any upcoming changes in pricing (which the company may announce ahead of time) may also trigger stockpiling by their customers, resulting in sudden jumps in demand. So, there can be both trend and seasonality factors

Their general manager needs to be able to predict demand for their products on a monthly, quarterly, and annual basis so that he can plan the production using their limited resources and his department's budget. He makes use of time series forecasting models to predict the potential demand for each of their product lines. By studying the seasonal patterns and growth trends, he can better prepare their production lines. For example, studying seasonality in the sales for the #2 wax tape, which is heavily used in cold climates, reveals that March and April are the months with the highest number of orders placed as customers buy them ahead of the maintenance seasons starting in the summer months. So the plant manager can dedicate most of their production lines to manufacturing the #2 tape during these months. This insight would not be known unless a time series analysis was performed.

(Kotu and Deshpande, 2019, p.399)

Time Series Forecasting
Trends and Seasonal Variations

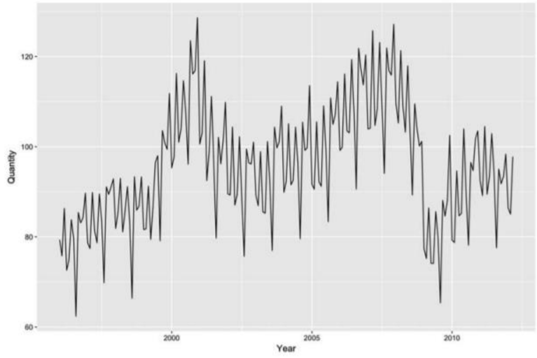


A time series analysis can reveal trends and seasonal patterns.

Time Series Forecasting Decomposition

Time series data capture multiple underlying phenomenon.

Below, the overall drug sales trends upward accelerating in the 2000s. There is yearly seasonality in the time series of drug sales with a spike in drug sales at the start of a year and a dip every February. The pattern in 2007 is odd when compared with prior years or 2008 and assumed as noise in the time series.



A time series analysis can reveal trends and seasonal patterns.

A Time Series Analysis can reveal trends and seasonal patterns (Kotu and Deshpande, 2019, p.400)

Time Series Forecasting Decomposition

Trend - long-term tendency of the data representing change from one period to next.

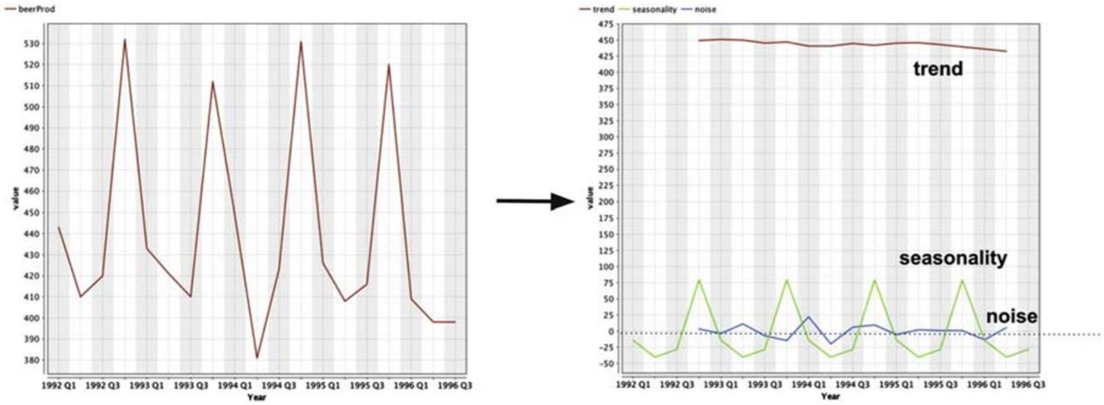
Seasonality - repetitive behaviour during a cycle of time, can be further split into hourly, daily, weekly, monthly, quarterly, and yearly seasonality.
For an online finance portal such as Yahoo Finance, visits show daily seasonality with a difference in the number of people accessing the portal or the app during the day (market open hours) and the night. Weekdays show higher traffic (and revenue) than during the weekends. Online advertising spend shows quarterly seasonality as the advertisers adjust marketing spend during the end of the quarter. Revenue shows yearly seasonality accounting for Christmas and holidays.

Cycle - longer-than-a-year patterns where there is no specific time frames between the cycles such as the economic cycle of booms and crashes.

Noise - anything not represented by level, trend, seasonality, or cyclic component. The noise component is unpredictable but follows normal distribution in ideal cases. All the time series datasets will have noise.

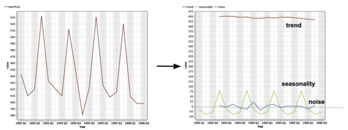
Time Series Forecasting Decomposition

In the time series decomposition can be classified into **additive decomposition** and **multiplicative decomposition**, based on the nature of the different components and how they are composed.



A Time Series Analysis can reveal trends and seasonal patterns (Kotu and Deshpande, 2019, p.400)

Time Series Forecasting Decomposition



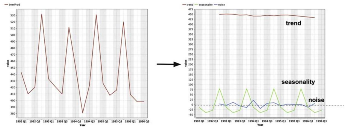
The trend and seasonality are the systematic components of a time series. Systematic components can be forecasted. It is impossible to forecast noise—the non-systematic component of a time series.

In the time series decomposition can be classified into additive decomposition and multiplicative decomposition, based on the nature of the different components and how they are composed.

In an additive decomposition, the components are decomposed in such a way that when they are added together, the original time series can be obtained. Time series = Trend + Seasonality + Noise

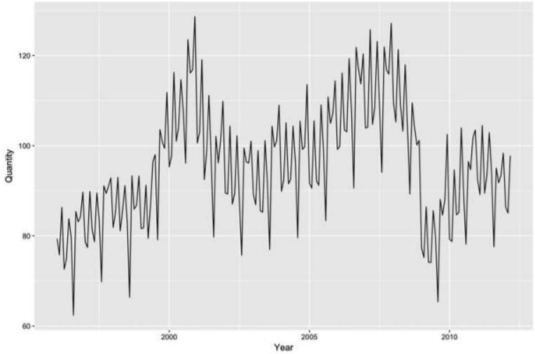
In the case of multiplicative decomposition the components are decomposed in the such a way that when they are multiplied together, the original time series can be derived back.
Time series = Trend X Seasonality X Noise

Time Series Forecasting Decomposition



If the magnitude of the seasonal fluctuation or the variation in trend changes with the level of the time series, then multiplicative time series decomposition is the better model.

The antidiabetic drug sales time series shows increasing seasonal fluctuation and exponential rise in the long-term trend. Hence, a multiplicative model is more appropriate.



A time series analysis can reveal trends and seasonal patterns.

A Time Series Analysis can reveal trends and seasonal patterns (Kotu and Deshpande, 2019, p.400)

Time Series Forecasting
Smoothing Based Methods

- **Time periods**: t = 1, 2, 3, …, n .

Seconds, days, weeks, months, or years.

- **Data series** $y_1, y_2, y_3, \dots, y_n$ corresponding to each time period above.

- **Forecasts**: F_{n+h} forecast for the h^{th} time period following.

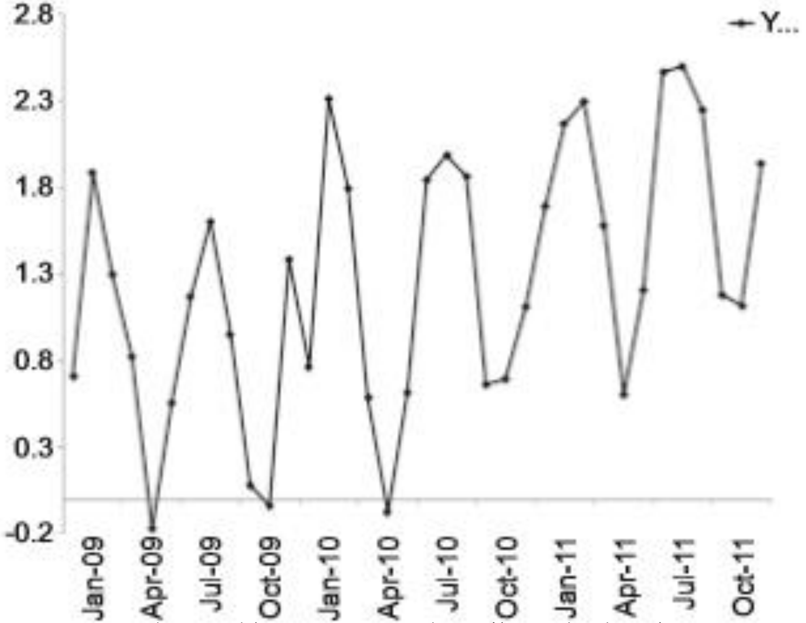
Usually $h = 1$, the next time period following the last data point. However h (also called the **horizon**) can be greater than 1 for some applications.

- **Forecast errors**: $e_t = y_t - F_t$ for any given time, t.

Time Series Forecasting
Smoothing Based Methods

Take a simple time series data function, $Y(t)$ where Y is the value of the time series at any time t and the data represent the value of Y over a 36-month period. $Y(t)$ is made up of a periodic (or seasonal) component and a random (or noise) component. Additionally, Y(t) has a small (in this case, upward) linear trend as well.

10_TimeS_10.2.1_simpleTS.csv



Teaching materials 'Data Science Concepts and Practice' (Kotu, Deshpande, 2019)

Time Series Forecasting
Smoothing Based Methods

- **Naïve Forecast**

Simplest forecasting "model." F_{n+1}, the forecast for the next period in the series, is given by the last data point of the series, y_n. $F_{n+1}(= y_{n+1}) = y_n$.

- **Simple Average**

Compute the next data point as an average of all the data points in the series. $F_{n+1} = AVERAGE(y_n\ , y_{n-1}, y_{n-2}, \dots, y_1)$.

- **Moving Average**

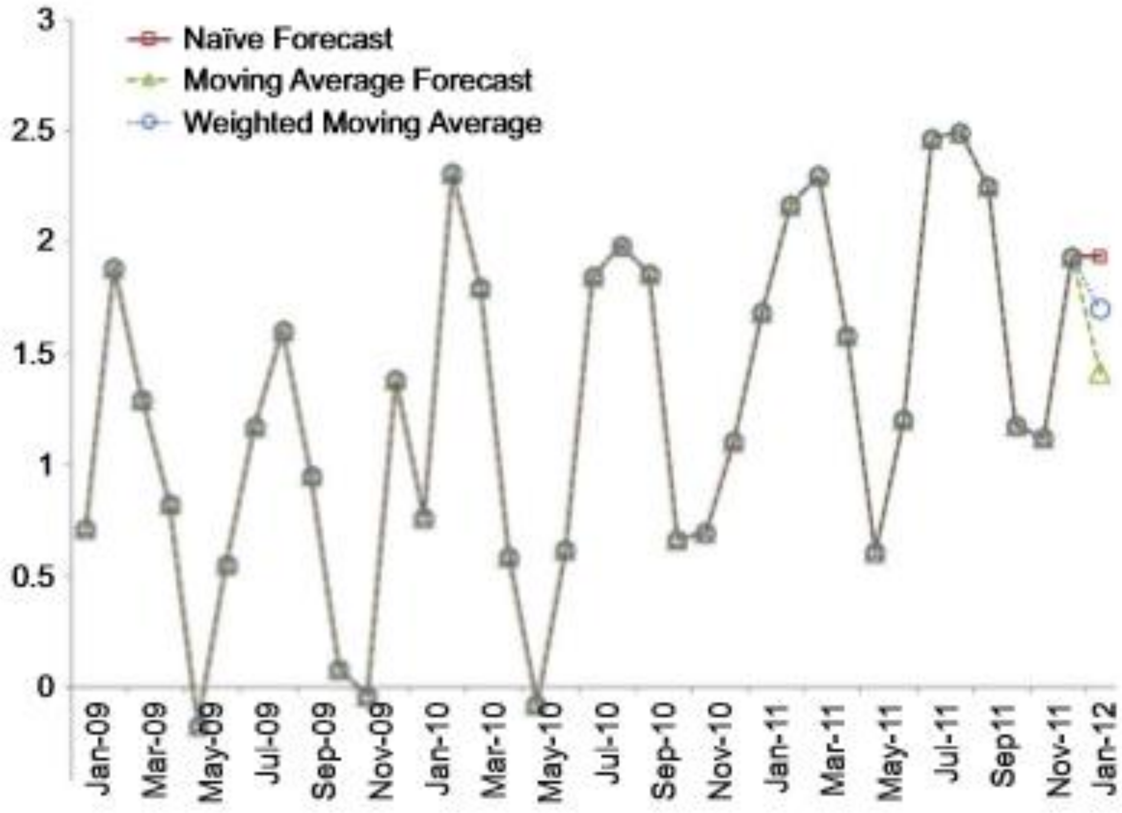
Select a window of the last "k" periods to calculate the average. The window for averaging keeps moving forward and thus returns a moving average. Suppose in our simple example that the window $k = 3$; then to predict the January data, we take a three-month average using the last three months.

- **Weighted Moving Average**

For some cases, the most recent value could have more influence than some of the earlier values. Most exponential growth occurs due to this simple effect. The forecast for the next period is given by the model

$$F_{n+1} = \frac{a * y_n + b * y_{n-1} + c * y_{n-2}}{a + b + c} \; where\ typically\ a > b > c$$

Time Series Forecasting
Smoothing Based Methods



Teaching materials 'Data Science Concepts and Practice' (Kotu, Deshpande, 2019)

Time Series Forecasting
Smoothing Based Methods

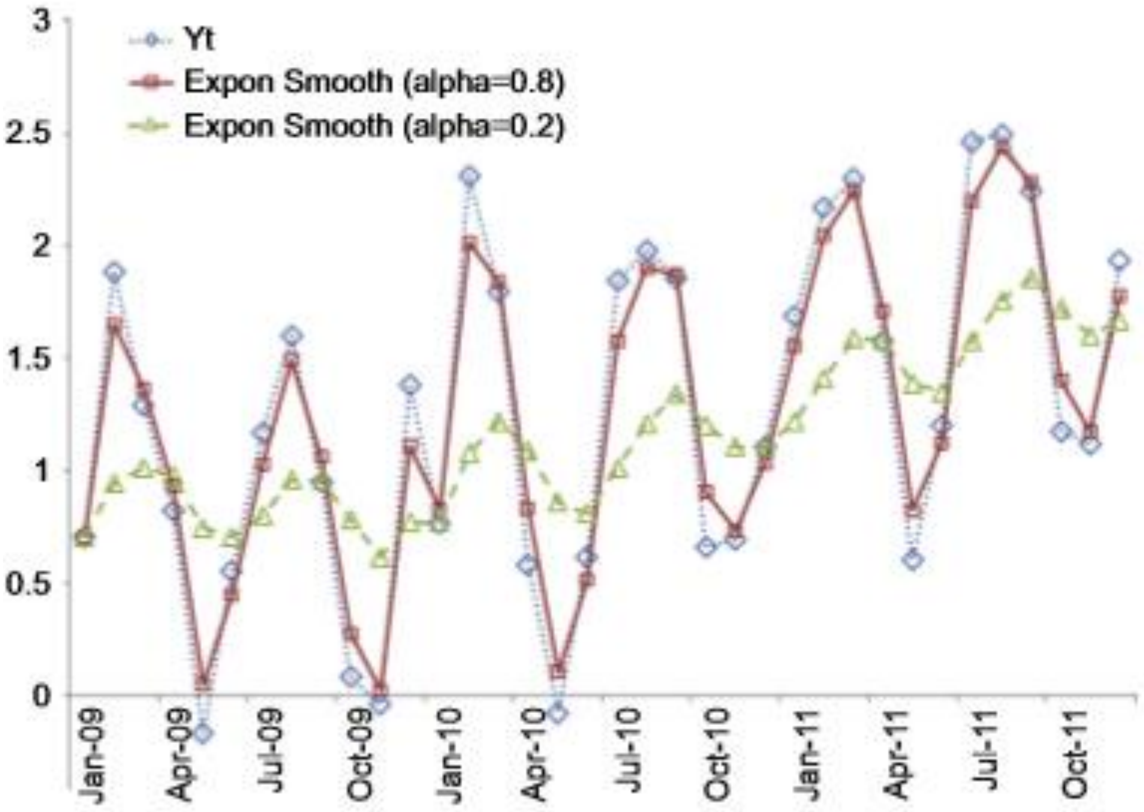
- **Exponential Smoothing (1956)**

Uses the previously forecast value for a given period to predict the value for the next period. For example making a February forecast using not only the actual January value but also the previously forecast January value. The new forecast can therefore "learn" the data a little better.

$$\alpha * \ y_n + (1-\alpha) * F_n \quad where\ \ 0 < \alpha\ < 1$$

If α is close to 1, then the previously forecast value of the last period has less weight than the actual value of the last period and vice versa. Note that α = 1 returns the naïve forecast putting more weight on actual values and a resulting curve that is closer to the actual curve, but using a lower α results in putting more emphasis on a previously forecast value and results in a smoother but less accurate fit. Typical values for α range from 0.2 to 0.4 in practice.

Time Series Forecasting
Smoothing Based Methods



Time Series Forecasting
Smoothing Based Methods

- **Exponential Smoothing cont.**

This simple exponential smoothing is the basis for a number of very common data-driven forecasting methods. However, forecasts can only be made one-step ahead by assuming $F_{n+h} = F_{n+1}$. To make longer horizon forecasts, $h \gg 1$, trend and seasonality information also need to be considered and the simple exponential smoothing methods quickly become more complicated.

A time series is made up of what is known as nonstationary data. “Nonstationary” means that the series typically demonstrates a trend and a seasonal pattern in addition to “normal” fluctuations. Most time series can be decomposed into components: trend, seasonality, and random noise as shown on next slide.

Once trend and seasonality can be captured, the value at any time in the future can be forecast, not just one step ahead values.

Time Series Forecasting
Smoothing Based Methods

- **Holt's Two-Parameter Exponential Smoothing**

Spreadsheets are commonly used to output trend lines on scatterplots. A trend is an averaged long-term tendency of a time series. The simplified exponential smoothing model described earlier is not very effective at capturing trends. An extension of this technique called Holt's two-parameter exponential smoothing is needed to accomplish this.

The previous exponential smoothing technique simply calculated the **average value** of the time series at $n+1$. If the series also has a trend, then an **average slope** of the series needs to be estimated as well.

Holt's two-parameter smoothing accomplishes this by inclusion of another parameter, β. A smoothing equation is constructed for the average trend at $n+1$. Using two parameters, α and β, any time series with a trend can be modelled and therefore forecast. The forecast can be expressed as a sum of the two components, average value or "level" of series, L_n, and trend, T_n,

$$F_{n+1} = L_n + T_n \; where$$

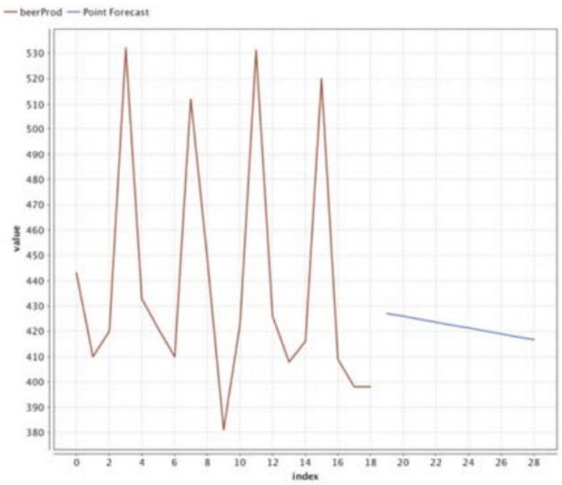
$$L_n = \alpha * y_n + (1-\alpha) * (L_{n-1} + T_{n-1}) \; and$$

$$T_n = \beta * (L_n - L_{n-1}) + (1-\beta) * T_{n-1}$$

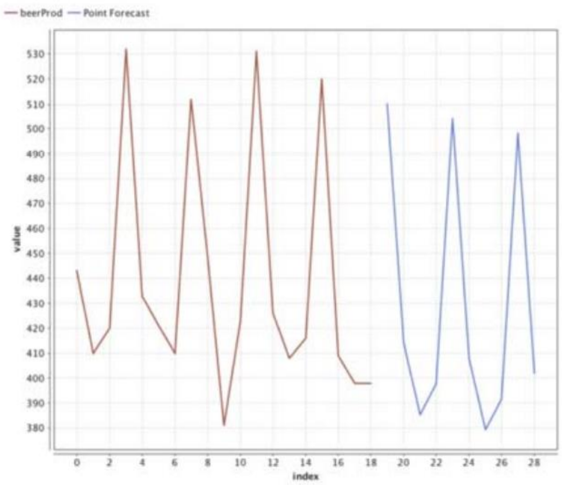
Time Series Forecasting
Smoothing Based Methods

- **Holt-Winters' Three-Parameter Exponential Smoothing**

When a time series contains seasonality in addition to a trend, another parameter, γ, is used to estimate the seasonal component of the time series.



Holt's Two-Parameter Smoothing



Holt-Winters' Three-Parameter Smoothing

Holt's and Holt's Winters Smoothing
(Kotu and Deshpande, 2019, p.414)

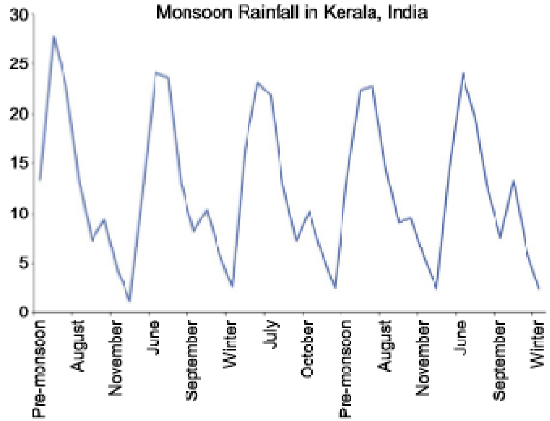
Time Series Forecasting
Regression Based Methods

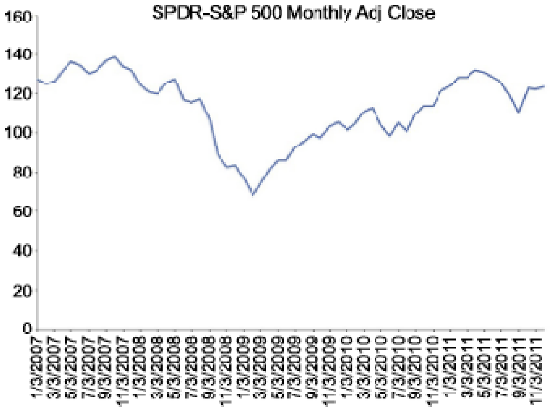
Model-driven approaches to time series forecasting overcome the one-step-ahead limitation of some of the data-driven methods. Time is the predictor or independent variable and the time series value is the dependent variable.

Model-driven methods are generally preferable when the time series appears to have a "global" pattern. The idea is that the model parameters will be able to capture these patterns and thus enable us to make predictions for any step ahead in the future under the assumption that this pattern is going to repeat.

For a time series with local patterns instead of a global pattern, using the model-driven approach requires specifying how and when the patterns change, which is difficult. For such a series, data-driven approaches work best because these methods usually rely on extrapolating the most recent local pattern as we saw earlier.

Time Series Forecasting
Model-driven vs Data-driven applications

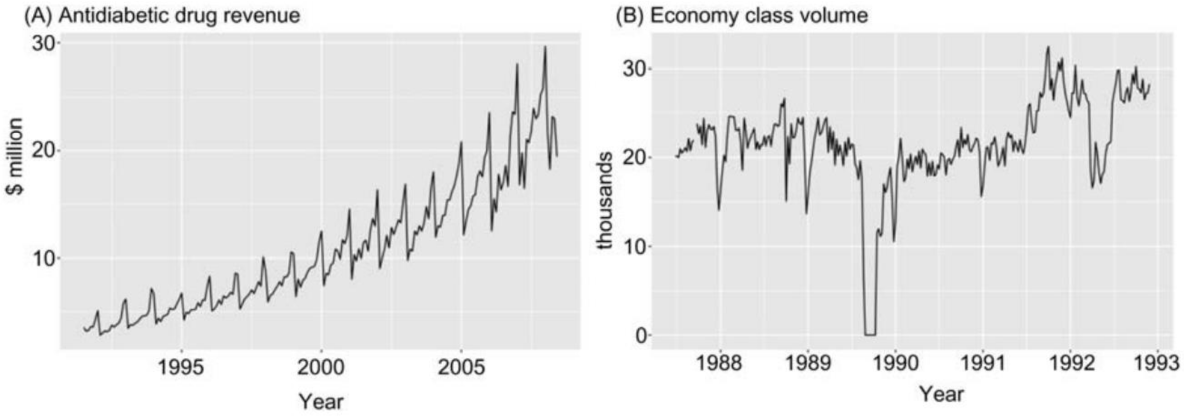




The following figures show annual monsoon precipitation in Southwest India averaged over a five-year period and the adjusted month-end closing prices of the SPDR S&P 500 (SPY) Index over another five-year period.

A model-driven forecasting method would work well for the rainfall series. However, the financial time series shows no clear start or end for any patterns. It is preferable to use data-driven methods to forecast this second series.

Time Series Forecasting
Model-driven vs Data-driven applications



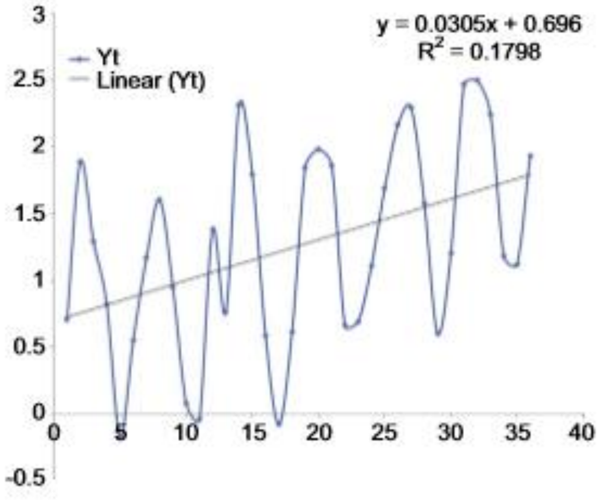
(Kotu and Deshpande, 2019, p.415)

Above two time series: (A) shows antidiabetic drug revenue and (B) shows the economy class passenger volume in Sydney-Melbourne route. A regression-based forecasting method would work well for the antidiabetic drug revenue series because it has a global pattern. However, the passenger volume series shows no clear start or end for any patterns. It is preferable to use smoothing based methods to attempt to forecast this second series.

Time Series Forecasting
Regression Based Methods

Linear Regression

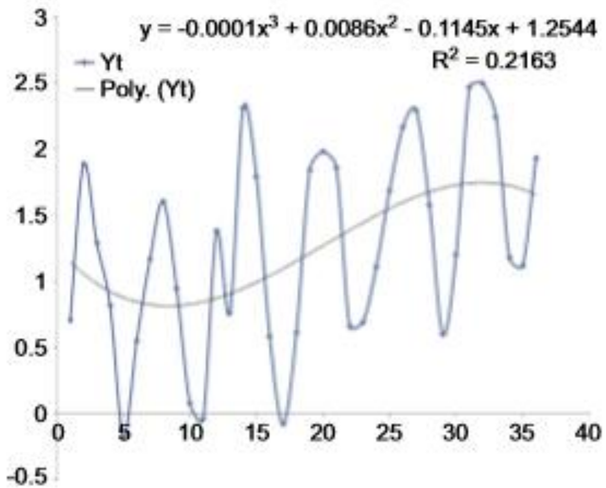
The simplest of the model-driven approaches. The time period is assumed as the independent predictor variable. The figure below shows a simple linear regression fitted to a 36-month dataset. The linear regression model is able to capture the long-term tendency of the series, but it does a very poor job of fitting the data. This is reflected in the R^2 value.



Time Series Forecasting
Regression Based Methods

Polynomial Regression

Polynomial regression is similar to linear regression except that higher-degree functions of the independent variable are used (squares and cubes). The figure below however shows that the polynomial doesn't do a significantly better job in fitting the data is reflected in the R^2 value. However, in either of the linear or polynomial regression cases, we are not limited to a one-step-ahead forecast of the simple smoothing (data-driven) methods.



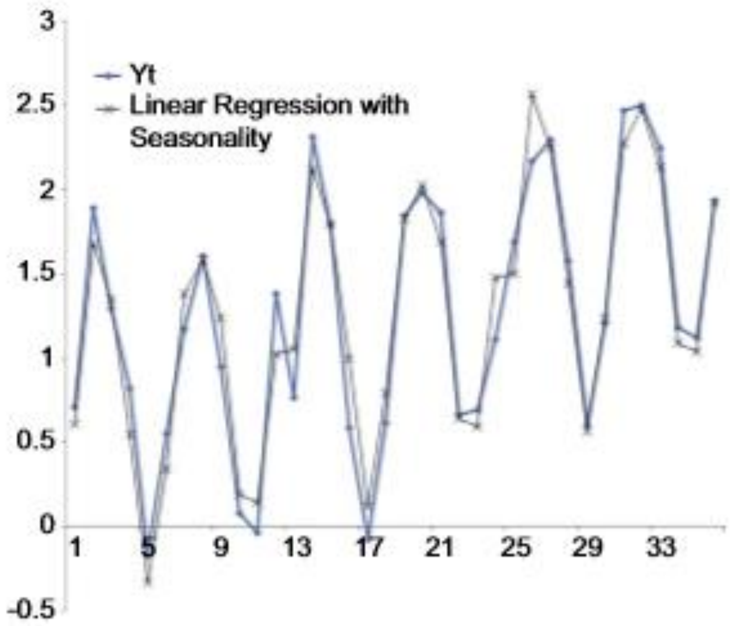
Linear Regression with Seasonality

The linear regression fit can be improved significantly by accounting for seasonality by introducing dummy variables triggering 1 or 0 for each month of the series. The linear regression model has only 13 variables, one for the time period and 12 dummy variables for each month of a year. The time-independent variable captures the trend and the 12 dummy variables capture seasonality. This regression equation can be used for predicting any future value beyond $n+1$ and has significantly more utility than the corresponding simpler counterparts in data-driven forecasting methods.

Seasonality Modeled via Linear Regression and the Accompanying Fit

Month	t	Dummy_1	Dummy_2	Dummy_3	Dummy_4	Dummy_5	Dummy_6	Dummy_7	Dummy_8	Dummy_9	Dummy_10	Dummy_11	Dummy_12	Yt
Jan	1	1	0	0	0	0	0	0	0	0	0	0	0	0.709
Feb	2	0	1	0	0	0	0	0	0	0	0	0	0	1.886
Mar	3	0	0	1	0	0	0	0	0	0	0	0	0	1.293
Apr	4	0	0	0	1	0	0	0	0	0	0	0	0	0.822
May	5	0	0	0	0	1	0	0	0	0	0	0	0	–0.173
Jun	6	0	0	0	0	0	1	0	0	0	0	0	0	0.552
Jul	7	0	0	0	0	0	0	1	0	0	0	0	0	1.169
Aug	8	0	0	0	0	0	0	0	1	0	0	0	0	1.604
Sep	9	0	0	0	0	0	0	0	0	1	0	0	0	0.949
Oct	10	0	0	0	0	0	0	0	0	0	1	0	0	0.08
Nov	11	0	0	0	0	0	0	0	0	0	0	1	0	–0.04
Dec	12	0	0	0	0	0	0	0	0	0	0	0	1	1.381

Time Series Forecasting
Regression Based Methods



Seasonality modeled via linear regression and the accompanying fit.

Time Series Forecasting
Regression Based Methods

ARIMA

Deshpande & Kotu (2019) ARIMA stands for Autoregressive Integrated Moving Average model and is one of the most popular models for time series forecasting.

- originally developed by Box and Jenkins in the 1970s.
- implementation relatively straightforward using statistical packages that support ARIMA functions.

Key concepts / building blocks: -

- autocorrelation,
- autoregression,
- stationary data,
- differentiation, and
- moving average of error.

Time Series Forecasting
Regression Based Methods

Autocorrelation

Correlation measures how two variables are dependent on each other or if they have a linear relationship with each other. Consider the time series shown in Figure.

Year	prod	prod-1	prod-2	prod-3	prod-4	prod-5	prod-6
1992 Q1	443	?	?	?	?	?	?
1992 Q2	410	443	?	?	?	?	?
1992 Q3	420	410	443	?	?	?	?
1992 Q4	532	420	410	443	?	?	?
1993 Q1	433	532	420	410	443	?	?
1993 Q2	421	433	532	420	410	443	?
1993 Q3	410	421	433	532	420	410	443
1993 Q4	512	410	421	433	532	420	410
1994 Q1	449	512	410	421	433	532	420
1994 Q2	381	449	512	410	421	433	532
1994 Q3	423	381	449	512	410	421	433
1994 Q4	531	423	381	449	512	410	421
1995 Q1	426	531	423	381	449	512	410

Lag Series and Auto Correlation (Kotu and Deshpande, 2019, p.420)

Time Series Forecasting
Regression Based Methods

Autocorrelation cont.

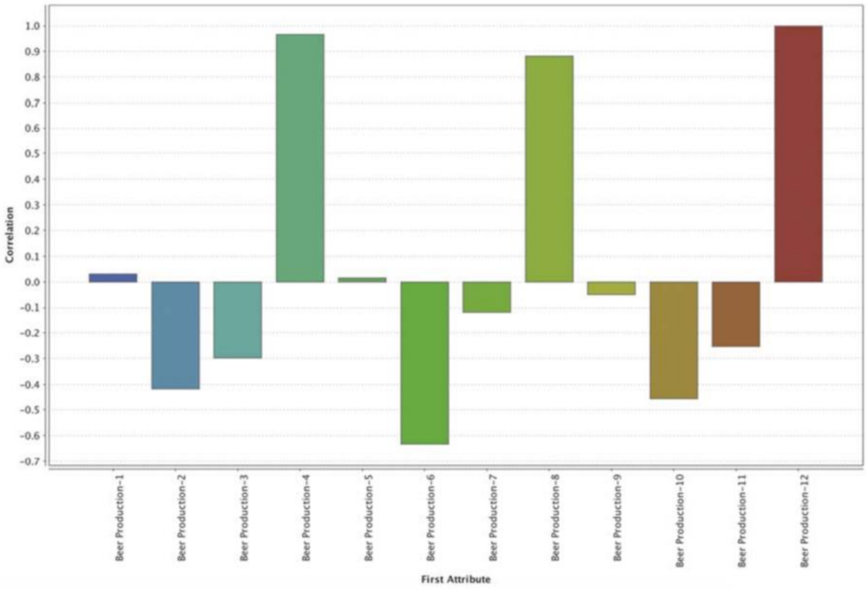
The second column “prod” shows the data for the simple time series. In the third column, data are lagged by one step. 1992 Q1 data is shown in 1992 Q2. This new series of values is termed a “1-lag” series. There are an additional 2-lag, 3-lag, ..., n-lag series in the dataset. Notice that there is a strong correlation between the original time series “prod” and 4-lag “prod-4.” They tend to move together. This phenomenon is called autocorrelation, where the time series is correlated with its own data points, with a lag.

As in a multivariate correlation matrix, one can measure the strength of correlation between the original time series and all the lag series. The plot of the resultant correlation matrix is called an Autocorrelation Function (ACF) chart. The ACF chart is used to study all the available seasonality in the time series.

Time Series Forecasting
Regression Based Methods

Autocorrelation cont.

From Figure it can be concluded that the time series is correlated with the 4th, 8th, and 12th lagged quarter due to the yearly seasonality. It is also evident than Q1 is negatively correlated with Q2 and Q3.



ACF Chart (Kotu and Deshpande, 2019, p.421)

Time Series Forecasting
Regression Based Methods

Autoregressive Models

Kotu & Deshpande (2019) Autoregressive models are regression models applied on lag series generated using the original time series. Recall in multiple linear regression, the output is a linear combination of multiple input variables. In the case of autoregression models, the output is the future datapoint and it can be expressed as a linear combination for past p data points. p is the lag window. The autoregressive model can be denoted as the equation:

$$y_t = l + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon$$

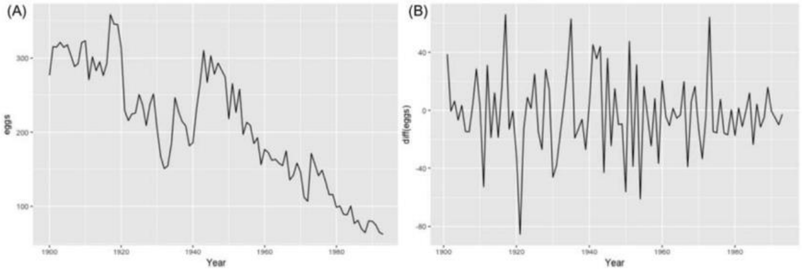
where, l is the level in the dataset and ε is the noise. α are the coefficients that need to be learned from the data. This can be referred to as an autoregressive model with p lags or an $AR(p)$ model. In an $AR(p)$ model, lag series is a new predictor used to fit the dependent variable, which is still the original series value, y_t.

Time Series Forecasting
Regression Based Methods

Stationary Data

Kotu & Deshpande (2019) In a time-series with trends or seasonality, the value is affected by time. A time series is called stationary when the value of time series is not dependent on time. For instance, **random white noise is a stationary time series**. Daily temperature at a location is not stationary as there will be a seasonal trend and it is affected by time. Meanwhile, the **noise component of a time series is stationary**.

Stationary time series do not have any means of being forecasted as they are completely random. Figure (A) is an example of nonstationary data because of the presence of both trend and seasonality and Figure (B) is an example of stationary data because there is no clear trend or seasonality.



(Kotu and Deshpande, 2019, p.422)

Time Series Forecasting
Regression Based Methods

Differencing

Kotu & Deshpane (201) A non-stationary time series can be converted to a stationary time series through a technique called differencing. Differencing series is the change between consecutive data points in the series. This is called first order differencing.

$$y'_t = y_t - y_{t-1}$$

Above shows a time series and a first order differenced time series. In some cases, just differencing once will still yield a nonstationary time series. In that case a second order differencing is required.

Second order differencing is the change between two consecutive data points in a first order differenced time series.

Differencing cont.

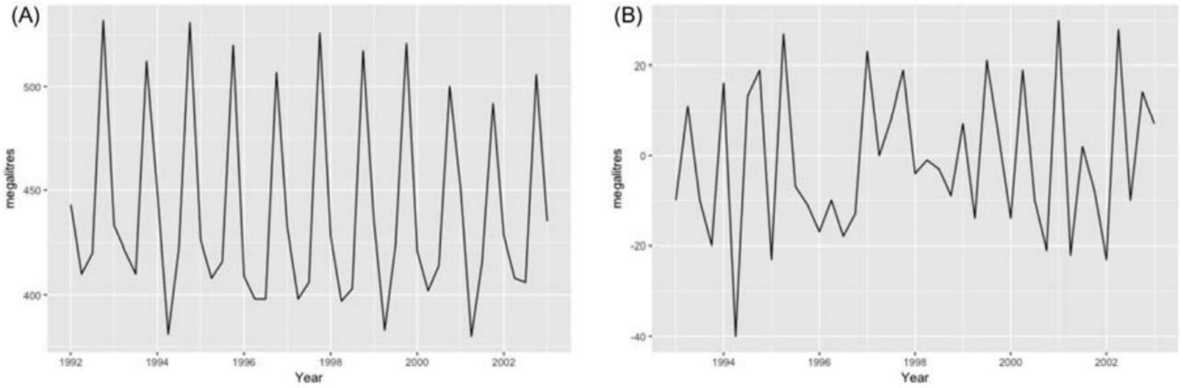
Differencing of order d is used to convert nonstationary time series to stationary time series. Seasonal differencing is the change between the same period in two different seasons. Assume a season has period, m.

$$y'_t = y_t - y_{t-m}$$

This is similar to the Year-over-Year metric used commonly in business financial reports. It is also called m-lag first order differencing.

Differencing cont.

Figure below shows the seasonal differencing of the Australian Beer production dataset and the seasonal first order differencing of the same series with the seasonal lag as 4—to factor in the number of quarters in a year.



(Kotu and Deshpande, 2019, p.423)

Time Series Forecasting
Regression Based Methods

Moving Average of Error

In addition to creating a regression of actual past “p” values as shown in one can also create a regression equation involving forecast errors of past data and use it as a predictor. Consider this equation with:

$$y_t = I + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

where e_i is the forecast error of data point i. This makes sense for the past data points but not for data point t because it is still being forecasted. Hence, e_t is assumed as white noise. The regression equation for y_t can be understood as the weighted (θ) moving average of past q forecast errors. This is called Moving Average with q lags model or $MA(q)$.

Time Series Forecasting
Regression Based Methods

Autoregressive Integrated Moving Average

The Autoregressive Integrated Moving Average (ARIMA) model is a combination of the differenced autoregressive model with the moving average model.

$$y'_t = I + \alpha_1 y'_{t-1} + \alpha_2 y'_{t-2} + \cdots + \alpha_p y'_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

AR - the time series is regressed on its own past data.
MA - the forecast error is a linear combination of past respective errors.
I - data values have been replaced with differenced values of d order to obtain stationary data.

The predictors are the lagged p data points for the autoregressive part and the lagged q errors are for the moving average part, which are all differenced.

Time Series Forecasting
Regression Based Methods

Kotu & Deshpande (2019) The prediction is the differenced y_t in the d^{th} order. This is called the ARIMA (p, d, q) model.

Estimating the coefficients α and θ for a given p, d, q is what ARIMA does when it learns from the training data in a time series. Specifying p, d, q can be tricky (and a key limitation) but one can try out different combinations and evaluate the performance of the model.

Once the ARIMA model is specified with the value of p, d, q the coefficients of the equation need to be estimated. The most common way to estimate is through the Maximum Likelihood Estimation. It is similar to the Least Square Estimation for the regression equation, except MLE finds the coefficients of the model in such a way that it maximizes the chances of finding the actual data.

Time Series Forecasting
Regression Based Methods

ARIMA is a generalized model. Some of the models discussed are special cases of an ARIMA model. For example,

- ARIMA $(0,1,0)$ is expressed as $y_t = y_{t-1} + \varepsilon$. It is the naive model with error, which is called the Random walk model.

- ARIMA $(0,1,0)$ is expressed as $y_t = y_{t-1} + \varepsilon + c$. It is a random walk model with a constant trend. It is called random walk with drift.

- ARIMA $(0,0,0)$ is $y_t = e$ or white noise

- ARIMA $(\mathrm{p}, 0,0)$ is the autoregressive model

Time Series Forecasting
Machine Learning Based Methods

Data Mining Learn Examples:-

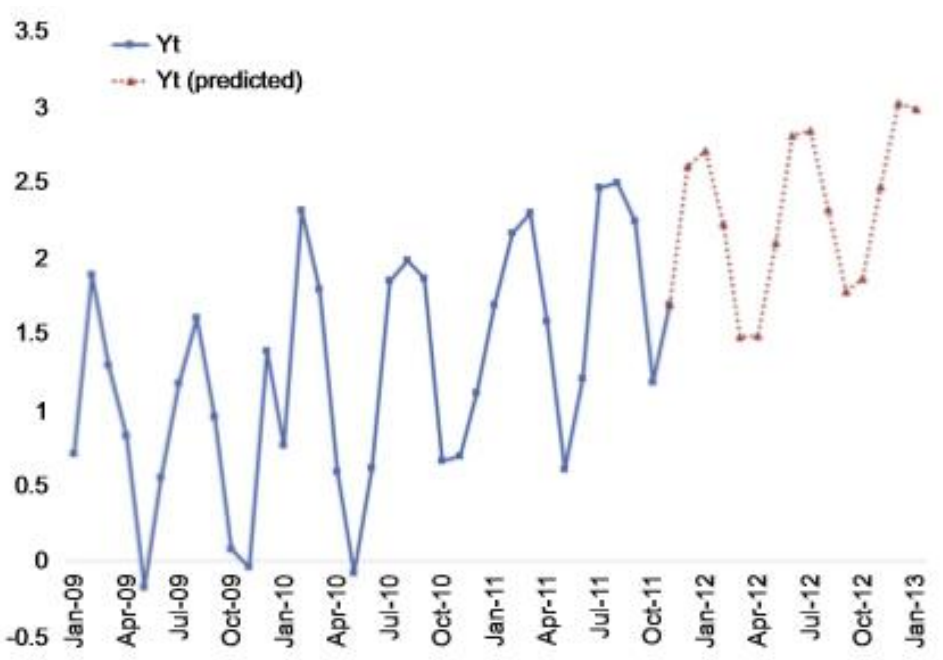
- 10_TimeS_10.2.1_simpleTS.csv
- 10_TimeS_10.2.2_simple_windowing_ts.rmp
- 10_TimeS_10.2.3_generate_forecasts_ts.rmp

Time Series Forecasting
Machine Learning Based Methods

Data Mining Learn Examples (using 10_TimeS_10.2.1_simpleTS.csv)

- 10_TimeS_10.2.2_simple_windowing_ts.rmp
- 10_TimeS_10.2.3_generate_forecasts_ts.rmp

	A	B
1	Date	Normalized Profits
2	01/01/2009	0.709
3	02/01/2009	1.886
4	03/01/2009	1.293
5	04/01/2009	0.822
6	05/01/2009	-0.173
7	06/01/2009	0.552
8	07/01/2009	1.169
9	08/01/2009	1.604
10	09/01/2009	0.949
11	10/01/2009	0.08
12	11/01/2009	-0.04
13	12/01/2009	1.381
32	07/01/2011	2.466
33	08/01/2011	2.497
34	09/01/2011	2.245
35	10/01/2011	1.179
36	11/01/2011	1.119
37	12/01/2011	1.934



Teaching materials ‘Data Science Concepts and Practice’ (Kotu, Deshpande, 2019)

Time Series Forecasting – Machine Learning Based Implementation Use Case cont.

Following Windowing transformation, Time Series Analysis can be performed by applying any of the available appropriate supervised learning algorithms to “predict” the label variable using the example set. The available Performance operators can also be used to assess the fitness of the learned model(s).

In summary the process consists of three steps:

(1) set up windowing;
(2) train the model with several different algorithms; and
(3) generate the forecasts.

Time Series Forecasting – Machine Learning Based Implementation : Step 1 Set Up Windowing

The process window in following slide shows the necessary operators for windowing.

1. The **Set Role** operator should set the **date** attribute as an **ID**. All time series will have a date column and this must be treated with special care.

2. The **Select Attributes** operator is used to select the attributes to forecast in this case “inputYt” is selected.

3. The **Windowing** operator. (Note:- the Series extension must be installed)

4. The **Filter Examples** operator may be used to remove attributes with missing values.

Time Series Forecasting – Machine Learning Based Implementation : Step 1 Set Up Windowing cont.

The main items to consider in Windowing are the following:

- **Window size**:

Determines how many "attributes" are created for the cross-sectional data. Each row of the original time series within the window width will become a new attribute. For this example set w = 6.

- **Step size**:

Determines how to advance the window. Set s = 1.

- **Horizon**:

Determines how far out to make the forecast. If the window size is 6 and the horizon is 1, then the seventh row of the original time series becomes the first sample for the "label" variable. Set h = 1.

Time Series Forecasting – Machine Learning Based Implementation : Step 1 Set Up Windowing cont.

Note:- for the window selected and shown in the box, the target or response variable value is the value from Jul 1, 2009. When training any algorithm using this data, the attributes labelled inputYt-5 through inputYt-0 form the independent variables.

Date	inputYt
Jan 1, 2009	0.709
Feb 1, 2009	1.886
Mar 1, 2009	1.293
Apr 1, 2009	0.822
May 1, 2009	-0.173
Jun 1, 2009	0.552
Jul 1, 2009	1.169
Aug 1, 2009	1.604
Sep 1, 2009	0.949
Oct 1, 2009	0.080
Nov 1, 2009	-0.040
Dec 1, 2009	1.381
Jan 1, 2010	0.761

Date	label	inputYt-5	inputYt-4	inputYt-3	inputYt-2	inputYt-1	inputYt-0
Jun 1, 2009	1.169	0.709	1.886	1.293	0.822	-0.173	0.552
Jul 1, 2009	1.604	1.886	1.293	0.822	-0.173	0.552	1.169
Aug 1, 2009	0.949	1.293	0.822	-0.173	0.552	1.169	1.604
Sep 1, 2009	0.080	0.822	-0.173	0.552	1.169	1.604	0.949
Oct 1, 2009	-0.040	-0.173	0.552	1.169	1.604	0.949	0.080
Nov 1, 2009	1.381	0.552	1.169	1.604	0.949	0.080	-0.040
Dec 1, 2009	0.761	1.169	1.604	0.949	0.080	-0.040	1.381
Jan 1, 2010	2.312	1.604	0.949	0.080	-0.040	1.381	0.761
Feb 1, 2010	1.795	0.949	0.080	-0.040	1.381	0.761	2.312
Mar 1, 2010	0.586	0.080	-0.040	1.381	0.761	2.312	1.795
Apr 1, 2010	-0.077	-0.040	1.381	0.761	2.312	1.795	0.586
May 1, 2010	0.613	1.381	0.761	2.312	1.795	0.586	-0.077

Window size = 6
Step size = 1
Horizon = 1

Using data from 6 rows (Jan 2009 – Jun 2009) of the window, a learner can be trained to predict the label which is the value of the time series in the next time step (Jul 2009) and so on.

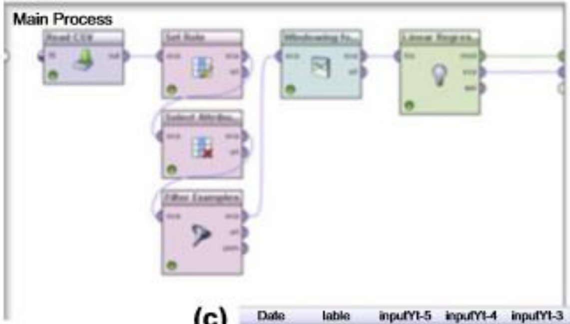
Output of windowing transformation.

Time Series Forecasting – Machine Learning Based Implementation : Step 2 Train the Model

Once the time series is encoded and transformed into a cross-sectional data set, any of the available machine learning algorithms such as regression, neural networks, or support vector machines, can be used to generate predictions. In this case we use linear regression to fit the “dependent” variable called label, given the “independent” variables inputYt-5 through inputYt-0.

Time Series Forecasting – Machine Learning Based Implementation : Step 2 Train the Model

(a)



(b)

Linear Regression

0.488 * inputYt-5
+ 0.286 * inputYt-4
+ 0.108 * inputYt-3
- 0.137 * inputYt-2
+ 0.292 * inputYt-0
+ 0.112

(c)

Date	lable	inputYt-5	inputYt-4	inputYt-3	inputYt-2	inputYt-1	inputYt-0
Mar 1, 2010	[illegible]	[illegible]	-0.040	1.381	0.761	2.312	1.795
Apr 1, 2010	-0.077	-0.040	1.381	0.761	2.312	1.795	0.585
May 1, 2010	0.613	1.381	0.761	2.312	1.795	0.585	-0.077
Jun 1, 2010	1.845	0.761	2.312	1.795	0.585	-0.077	0.613
Jul 1, 2010	1.984	2.312	1.795	0.585	-0.077	0.613	1.845
Aug 1, 2010	1.851	1.795	0.585	-0.077	0.613	1.845	1.984
Sep 1, 2010	0.651	0.586	-0.077	0.613	1.845	1.984	1.861
Oct 1, 2010	0.692	-0.077	0.613	1.845	1.984	1.861	0.661
Nov 1, 2010	1.108	0.613	1.845	1.984	1.861	0.661	0.692
Dec 1, 2010	1.688	1.845	1.984	1.861	0.661	0.692	1.108
Jan 1, 2011	2.157	1.984	1.861	0.661	0.692	1.108	1.688
Feb 1, 2011	2.285	1.861	0.661	0.692	1.108	1.688	2.167
Mar 1, 2011	1.577	0.661	0.692	1.108	1.688	2.167	2.295
Apr 1, 2011	0.601	0.692	1.108	1.688	2.167	2.295	1.577
May 1, 2011	1.201	1.108	1.688	2.167	2.295	1.577	0.601
Jun 1, 2011	2.456	1.688	2.167	2.295	1.577	0.601	1.201
Jul 1, 2011	2.497	2.167	2.295	1.577	0.601	1.201	2.455
Aug 1, 2011	2.245	2.295	1.577	0.601	1.201	2.455	2.497
Sep 1, 2011	1.179	1.577	0.601	1.201	2.455	2.497	2.245
Oct 1, 2011	1.119	0.601	1.201	2.455	2.497	2.245	1.179
Nov 1, 2011	1.934	1.201	2.455	2.497	2.245	1.179	1.119

Note: Using the process shown (a), the “label” variable is fitted using the six dependent variables via linear regression (b). Note that the label for any given row is the inputYt-0 for the next row (c).

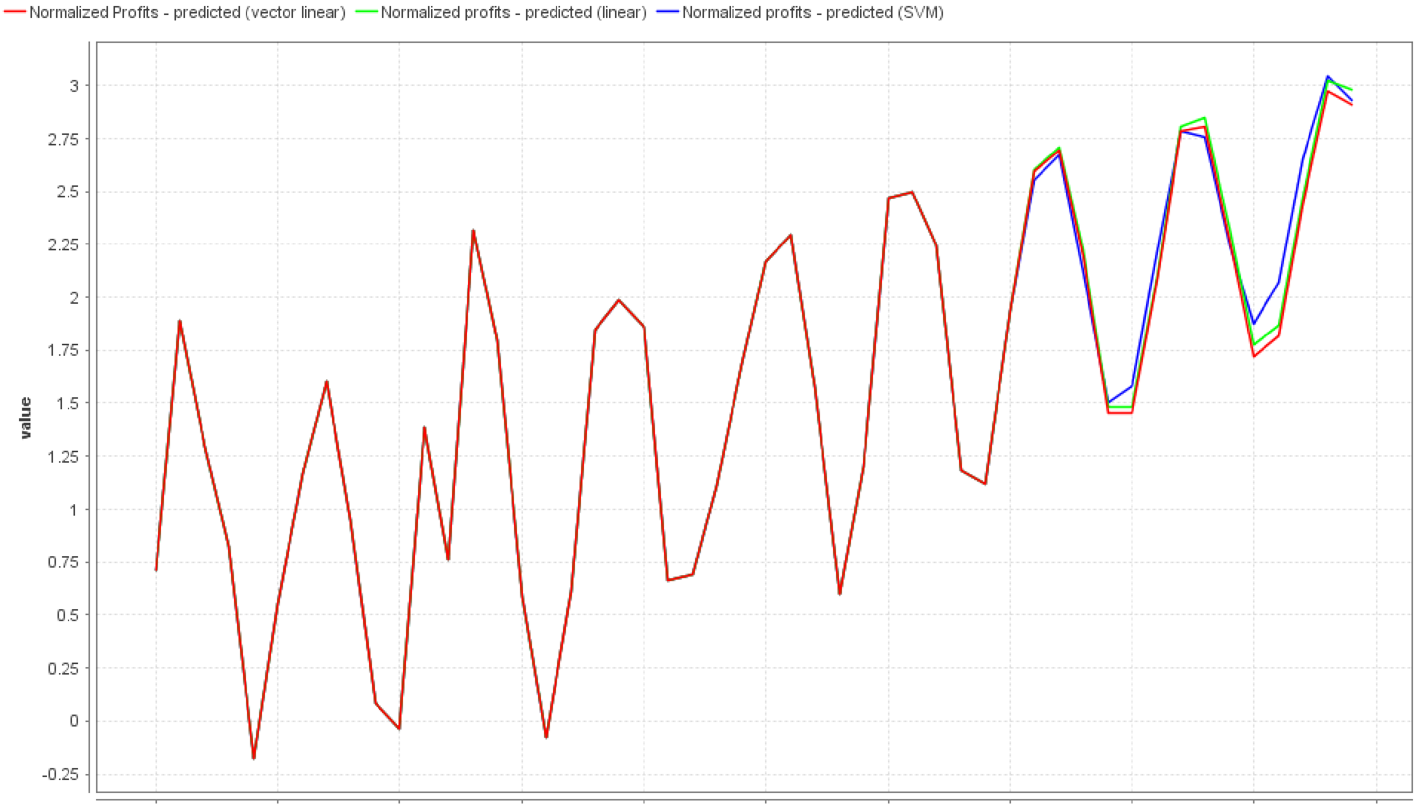
Time Series Forecasting – Machine Learning Based Implementation : Step 3 Prep. Gen. Forecasts

Once the model fitting is done, the next step is to start the forecasting process. Note that given this configuration of window size and horizon, only the forecast for the next step can be made.

In the example, the last row of the transformed data set corresponds to Nov. 1, 2011. The independent variables are values from June through November 2011 and the target or label variable is from December 2011. The regression equation and the values from the Nov. 1, 2011 row can be used to generate the forecast for January 2012.

The next row of data should run from August–January to predict February using the regression equation. The (actual) data from August to December and the predicted value for January can be used. Now the actual data from September–December plus predicted January and February values can be used to forecast March.

Time Series Forecasting – Machine Learning Based Implementation : Step 3 Generate Forecasts cont.



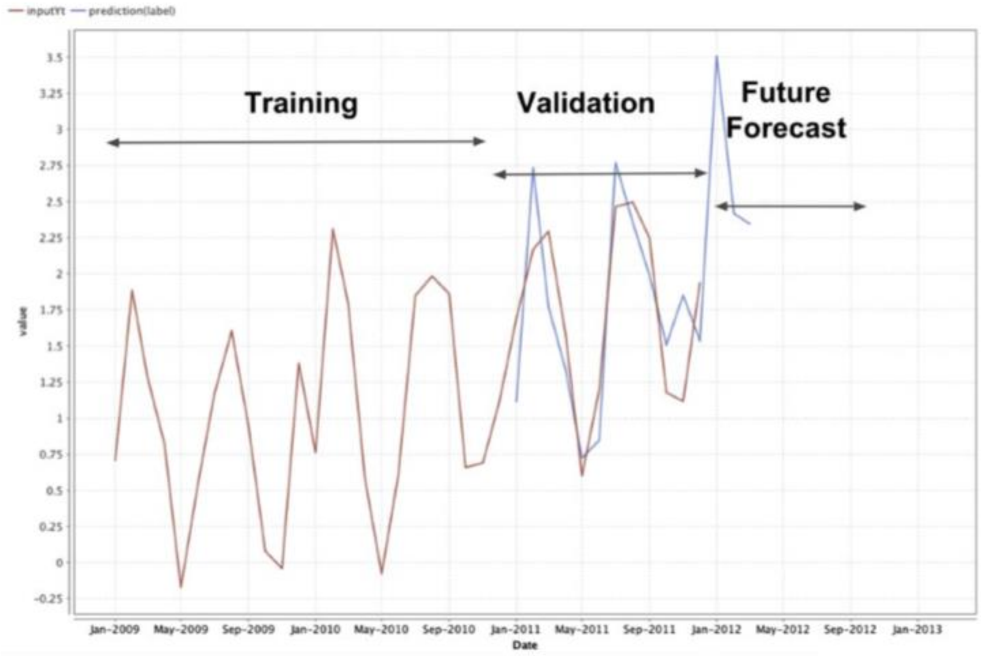
Time Series Forecasting
Measuring Performance

Validation Dataset

One way to measure accuracy is to measure the actual value post-fact and compare against the forecast and calculate the error. However, one would have to wait for the time to pass and the actual data to be measured. Residuals are calculated when a forecast model is fitted on training data. By all means, the model might have overfitted the training data and perform poorly with unseen future forecasts. For this reason, training residuals are not a good way to measure how the forecast models are going to perform in the real world.

Recall that in the case of the supervised learners, a set of known data is reserved for model validation. Similarly, some data points can be reserved in the time series as a validation dataset just to test the accuracy of the model. The training process uses data by restricting it to a point and the rest of the later time series is used for validation. As the model has not seen the validation dataset, deviation of actuals from the forecast is the forecast error of the model. The forecast error is the difference between the actual value $\boldsymbol{y_i}$ and the forecasted value $\boldsymbol{\hat{y}_i}$. The error measured for each of the data points can be aggregated to one metric to indicate the error of the forecasting model.

Time Series Forecasting
Measuring Performance



Validation dataset.

Forecasted Time Series (Kotu and Deshpande, 2019, p.437)

Time Series Forecasting
Measuring Performance

Some of the commonly used forecast accuracy aggregate metrics are:

Mean Absolute Error
The error of the individual data point may be positive or negative and may cancel each other out. To derive the overall forecast for the model, calculate the absolute error to aggregate all the residuals and average it. MAE is a simple metric and it is scale dependent. It is convenient to communicate the error of the revenue forecasting model as, for example, $\pm$\$900,000 per day.

Root Mean Squared Error
In some cases it is advantageous to penalize the individual point error with higher residuals. Even though two models have the same MAE, one might have consistent error and the other might have low errors for some points and high error for other points. RMSE penalizes the latter.

Cont.

Time Series Forecasting
Measuring Performance

Mean Absolute Percentage Error

Percentage error of a data point is a scale independent error that can be aggregated to form mean absolute percentage error. MAPE is useful to compare against multiple models across the different forecasting applications. For example, the quarterly revenue forecast, measured in USD, for a car brand might be $\pm$5% and the forecast for world-wide car demand, measured in quantity, might be $\pm$3%. The firm's ability to forecast the car demand is higher than the revenue forecast for one brand. Even though MAPE is easy to understand and scale independent, MAPE has a significant limitation when it applies to intermittent data where zero values are possible in actual time series. For example, profit or defects in a product. Zero value in the time series yields an infinite error rate (if the forecast is nonzero) and skews the result. MAPE is also meaningless when the zero point is not defined or arbitrarily defined, as in non-kelvin temperature scales.

Cont.

Time Series Forecasting
Measuring Performance

Mean Absolute Scaled Error

MASE is scale independent and overcomes the key limitations of MAPE by comparing the forecast values against a naive forecast. Scaled error is less than one if the forecast is better than naive forecast and greater than one if it is worse than naïve. One would want a scaled error much less than one for a good forecasting model.

It is the mean absolute error of the forecast values, divided by the mean absolute error of the in-sample one-step naive forecast. It was proposed in 2005 by statistician Rob J. Hyndman and Professor of Decision Sciences Anne B. Koehler, who described it as a "generally applicable measurement of forecast accuracy without the problems seen in the other measurements. The mean absolute scaled error has favourable properties when compared to other methods for calculating forecast errors, such as RMSE, and is therefore recommended for determining comparative accuracy of forecasts.

Time Series Forecasting
Discussion Points

Time series forecasting is one of the most widely used analytical applications in business and organizations. All organizations are forward looking and want to plan for the future.

Univariate time series forecasting treats prediction, essentially, as a single variable problem, whereas, multivariate time series may use many time concurred value series for prediction. If one has a series of points spaced over time, conventional forecasting uses smoothing and averaging to predict where the next few points will likely be. However, for complex systems such as the economy or demand of a product, point forecasts are unreliable because these systems are functions of hundreds if not thousands of variables. What is more valuable or useful is the ability to predict trends, rather than point forecasts. Trends can be predicted with greater confidence and reliability (i.e., Are the quantities going to trend up or down?), rather than the values or levels of these quantities. For this reason, using an ensemble of different modelling schemes such as artificial neural networks or support vector machines or polynomial regression can sometimes give highly accurate trend forecasts.

Time Series Forecasting
Best Practices

- **Understand the metric**: Investigate how the time series metric is derived. Is the metric influenced by other metrics or phenomenon that can be better candidates for the forecasting? For example, instead of forecasting profit, both revenue and cost can be forecasted, and profit can be calculated. This is particularly suitable when profit margins are low and can go back and forth between positive and negative values (loss).

- **Plot the time series**: A simple time series line chart reveals a wealth of information about the metric being investigated. Does the time series have a seasonal pattern? Long-term trends? Are the seasonality and trend linear or exponential? Is the series stationary? If the trend is exponential, can one derive log() series? Aggregate daily data to weeks and months to see the normalized trends.

- **Is it forecastable**: Check if the time series is forecastable using stationary checks (ARIMA - Stationary time series is when the mean and variance are constant over time. It is easier to predict when the series is stationary. Differencing is a method of transforming a non-stationary time series into a stationary one).

Time Series Forecasting
Best Practices

- **Decompose**: Identify trends and seasonality using decomposition methods. These techniques show how the time series can be split into multiple meaningful components.

- **Try them all:** Try several different methods mentioned in the forecasting taxonomy. For each method, perform residual checks using MAE or MAPE metric, evaluate forecasts using the validation period, select the best performing method and parameters using optimization functions, update the model using the full dataset for future forecasts.

- **Maintain the models:** Review models on a regular basis. Time series forecast models have a limited shelf life. Apart from feeding the latest data to the model, the model should be refreshed to make it relevant for the latest data. Building a model daily is not uncommon.

References

- Han, J., Pei, J., Tong, H. (2022) *Data Mining Concepts and Techniques.* 4th edn. Burlington, MA: Morgan Kaufmann Publishers.
- Hilpesh, Y., J. (2019) Python for Finance 2nd ed. O'Reilly Media Inc.
- Kotu, V. and Deshpande, B. (2019) *Data Science Concepts and Practice.* 2nd end. Burlington, MA: Morgan Kaufmann Publishers.

Base Academic Papers

- Box, G. A. (1970). Time series analysis: Forecasting and control. San Francisco, CA: Holding Day.
- Box, G. J. (2008). Time series analysis: Forecasting and control. Wiley Series in Probability and Statistics.
- Brown, R. G. (1956). Exponential smoothing for predicting demand. Cambridge, MA: Arthur D. Little.
- Hyndman R.A. (2014). Forecasting: Principles and practice. ,Otexts.org..
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice. 2nd edition. ,Otexts.org..
- Hyndman, Rob J., & Koehler, Anne B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22, 679688.

Base Academic Papers

- Shmueli G. (2011). Practical time series forecasting: A hands on guide. ,statistics.com..
- Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324342.