Time series Problems

Friday, January 19, 2024 11:35 AM

Very important notebook by daniel: https://dev.mrdbourke.com/tensorflow-deep-learning/02_neural_network_classification_in_tensorflow/#improving-a-model

https://www.uber.com/en-DE/blog/forecasting-introduction/

This particular links tell about the how Uber tries to predict forecasting and what they have used for market demand

What can we predict into the future; because prediction can always go wrong (https://otexts.com/fpp3/) this books tell about forecasting principles and fundamentals go through it.

https://arxiv.org/abs/1905.10437 (Nbeats model)

1. Horizon: Think of the horizon as the future you're looking at. If you're trying to predict the weather for the next seven days, your horizon is seven days. It's like gazing into the future to see how things will unfold.
2. Window: Now, imagine you have a window that you can slide over your historical data. This window helps you see what's been happening in the past. For instance, if you want to predict tomorrow's weather, you might look at the weather patterns in the past month or year. The window is like a tool that helps you focus on a specific period in the past to make your predictions for the future.

From <https://chat.openai.com/c/3b64a1495-0afe-468d-8276-c8b8d94e4feb>

Time series models :

Moving average https://machinelearningmastery.com/moving-average-smoothing-for-time-series-forecasting-python/

ARIMA (Autoregression Integrated Moving Average) https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/

sktime (Scikit-Learn for time series) https://github.com/alan-turing-institute/sktime

TensorFlow Decision Forests (random forest, gradient boosting trees) https://www.tensorflow.org/decision_forests

Facebook Kats (purpose-built forecasting and time series analysis library by Facebook) analysis library by Facebook) https://github.com/facebookresearch/Kats

LinkedIn Greykite (flexible, intuitive and fast forecasts) https://github.com/linkedin/greykite

To know about auto correlation and how it benefits naive model or forcasting model :(simple model). https://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424

- tf.keras.callbacks.EarlyStopping() - stop the model from training if it doesn't improve validation loss for 200 epochs and restore the best performing weights using restore_best_weights=True (this'll prevent the model from training for loooongggggg period of time without improvement)
- tf.keras.callbacks.ReduceLROnPlateau() - if the model's validation loss doesn't improve for 100 epochs, reduce the learning rate by 10x to try and help it make incremental improvements (the smaller the learning rate, the smaller updates a model tries to make)

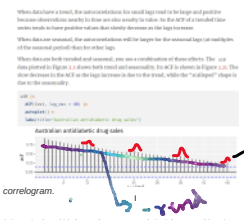
Some important points form the book Forecasting >Principles and Practice

Lag plots are a type of statistical visualization used in time series analysis to examine the relationship between a variable and its lagged values (past observations). These plots help assess autocorrelation, which is the correlation of a time series with its own past values. The basic idea is to scatter plot each data point against its lagged values to identify patterns or dependencies.

1. Axis Values:
   - The x-axis represents the current observations (or time points) of the time series.
   - The y-axis represents the lagged observations, usually with a lag of 1 (previous value), 2 (value two time points ago), and so on.
2. Scatter Plot:
   - Each point on the lag plot corresponds to a pair of observations, where one is the current value, and the other is its lagged value.
   - The scatter plot visually displays how closely related the current observation is to its lagged values.
3. Interpretation:
   - If there is a strong correlation between the current observation and its lagged values, you will observe a recognizable pattern or trend in the lag plot.
   - Common patterns include diagonal lines, curves, or clusters of points, indicating a relationship between the variable and its past values.
4. Autocorrelation Assessment:
   - Lag plots are particularly useful for identifying autocorrelation in time series data. Autocorrelation is the correlation of a signal with a delayed copy of itself. If there is a strong autocorrelation, the lag plot will reveal a structured pattern.
5. Randomness Check:
   - In contrast, a lag plot with no apparent pattern suggests randomness or lack of autocorrelation in the time series data.

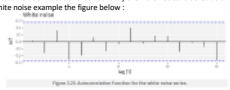
correlation measures the extent of a linear relationship

Autocorrelation



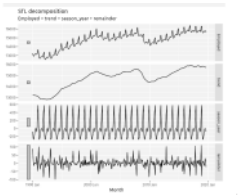
correlogram.

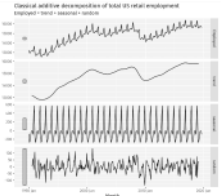
A time series data which dose not have a auto correlation is known as white noise. for white noise we consider that each autocorrelation is equal to zero ,but we know that the relation cannot be exactly zero , so we consider a boundary of 95% that is if the autocorrealtion is between that boundary then it is a white noise ,if one or 5 % of the autocorrelation is out of that boundary then it is not considered as a white noise example the figure below :

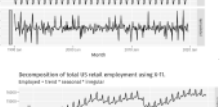


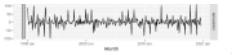
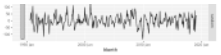
Additive and multiplicative decomposition

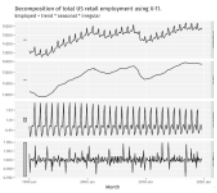
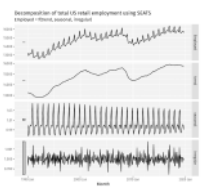
TIME SERIES DECOMPOSITION







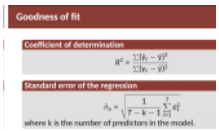



Residuals or difference between the predicted values and actual values:

If a transformation has been used in the model, then it is often useful to look at residuals on the transformed scale. We call these "**innovation residuals**"

Useful properties of residuals is that:
They have constant variance and they have normal distribution; This does not mean that all the residuals should display this property ,bit it is useful to have. And makes life easier

Linear regression.



Why there is necessary that we consider intercept for liner regression?

In layman terms, adding an intercept to a linear regression model is like including a baseline value that represents the starting point or average level of the dependent variable when all the independent variables are set to zero.

Imagine you are trying to predict someone's weight based on their height. The linear regression equation might look like this:

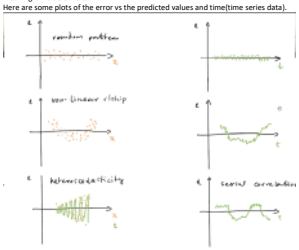
\[\text{Weight} = \text{Intercept} + (\text{Coefficient} \times \text{Height}) \]

Here:
- The \(\text{Intercept}\) is a constant term, and it represents the expected weight when the person's height is zero. In reality, height can't be zero, but the intercept helps capture the baseline weight that is not influenced by height.

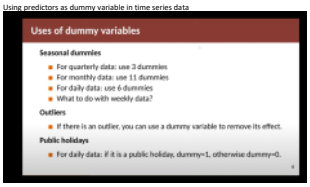
- \(\text{Coefficient}\) is the slope of the line, indicating how much the weight changes for each unit increase in height.

Without the intercept, the line would have to pass through the origin (0,0), and the model would assume that when height is zero, weight is also zero. This might not make sense in many real-world scenarios. The intercept allows the line to start at a more meaningful point, reflecting the baseline value when the independent variables are all zero or not applicable.

So, the intercept provides a starting point for the relationship between the variables and allows the model to better capture the real-world situation where some variables may not be relevant or have a meaningful value of zero.
Here are some plots of the error vs the predicted values and time(time series data).



Using predictors as dummy variable in time series data



How to select the best predictors to use in regression models.

Adjusted R squared:
The normal R2 tells us how well a model fits to historical data ,but not how well the model forecast in future.
And there is no degree of freedom, adding a variable will increase the value of R2,for these reasons R2 should be used to determine whether the model will give good predictions.

An alternative which is designed to overcome these problems is the adjusted $R2$ (also called "R-bar-squared"): $\overline{R}2=1-(1-R2)T-1T-k-1,$

From <https://otexts.com/fpp3/selecting-predictors.html>

BIC :

Which measure should we use?

While R^2 is widely used, and has been around longer than the other measures, its tendency to select too many predictor variables makes it less suitable for forecasting.

Many statisticians like to use the BIC because it has the feature that if there is a true underlying model, the BIC will select that model given enough data. However, in reality, there is rarely, if ever, a true underlying model, and even if there was a true underlying model, selecting that model will not necessarily give the best forecasts (because the parameter estimates may not be accurate).

Consequently, we recommend that one of the AICc, AIC, or CV statistics be used, each of which has forecasting as their objective. If the value of T is large enough, they will all lead to the same model. In most of the examples in this book, we use the AICc value to select the forecasting model.

Checking different predicters which has the lowest AIC

Imagine that you have 40 predictors and you can fit 2 ^ 40 models which is tomuch so a method of forward stepwise regression can be used, where you start with only the intercept and then keep on adding each one of the predictors and see whether the accuracy is improving this is done till the model show no improvement.

1. Ex-Ante Forecasts:
   - Layman's Explanation: Ex-ante forecasts are predictions or estimates made before an event or period has occurred. It's like trying to guess the outcome of a game before it starts. For example, if you predict the score of a soccer match before the game begins, that's an ex-ante forecast.
2. Ex-Post Forecasts:
   - Layman's Explanation: Ex-post forecasts, on the other hand, are predictions made after the event or period has already taken place. It's like making a guess about the score of a soccer match after it has ended. If you predict the score once the game is over, that's an ex-post forecast.

A comparative evaluation of ex-ante forecasts and ex-post forecasts can help to separate out the sources of forecast uncertainty. This will show whether forecast errors have arisen due to poor forecasts of the predictor or due to a poor forecasting model.4

NON LINEAR TRENDS

The simplest way of modelling a non linear relationship is to transform either the forecast variable or the predictor variable. The most commonly used variable is log functions.

A log-log functional form is specified as

$$\log y = \beta_0 + \beta_1 \log x + \varepsilon.$$

In this model, the slope β_1 can be interpreted as an elasticity: a 1% increase in x gives a β_1% average change in y (rounding from a 0% increase in x. Other transformations can also be specified. The log–linear form is specified by only transforming the forecast variable and the linear–log form is obtained by transforming the predictor.

The sentence is referring to the interpretation of the coefficient \(\beta_1 \) in a linear regression model, specifically in the context of elasticity.

- **Slope \(\beta_1 \):** This is a coefficient in the linear regression equation that represents the change in the dependent variable (\(y\)) for a one-unit change in the independent variable.

- **Interpretation as Elasticity:** When the sentence mentions that \(\beta_1 \) can be interpreted as an elasticity, it means that \(\beta_1 \) represents the average percentage change in the dependent variable (\(y\)) resulting from a 1% increase in the independent variable.

- **Elasticity Concept:** Elasticity is a measure of the responsiveness of one variable to changes in another variable. In this case, it's the responsiveness of \(y\) to changes in the independent variable.

- **Example Interpretation:** If \(\beta_1 = 0.05 \), it suggests that, on average, a 1% increase in the independent variable is associated with a 0.05% increase in the dependent variable \(y\).

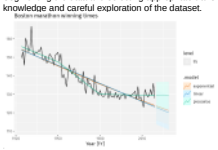
This interpretation is useful when dealing with percentage changes, especially in situations where the relationship between variables is better understood in terms of proportional or percentage adjustments rather than absolute changes.

Piecewise transformation

When log transformation is not suitable or does not adequately address the characteristics of the data, piecewise transformations may be considered. A piecewise transformation involves dividing the data into different segments and applying distinct transformations to each segment.
Here's what it shows:

3. Nonlinearity: Piecewise transformations allow for capturing non-linear relationships in different regions of the data. Instead of assuming a single transformation applies uniformly across the entire range of the variable, piecewise transformations recognize that different parts of the data may follow different functional forms.
4. Adaptability: Piecewise transformations are more flexible and adaptable to complex data patterns. They can better handle situations where the relationship between variables changes across different ranges or exhibits multiple patterns.
5. Improved Fit: By applying different transformations to different segments, piecewise transformations aim to provide a better fit to the underlying data structure. This can lead to improved model accuracy and a more accurate representation of the relationships between variables.
6. Interpretability: Depending on the nature of the data, piecewise transformations may also offer interpretability advantages. Instead of applying a single, global transformation, understanding how the transformation varies across different segments can provide insights into the behavior of the variables.
7. Local Adjustment: Piecewise transformations enable local adjustments to be made where they are most needed. If certain parts of the data exhibit different characteristics or patterns, piecewise transformations allow tailored transformations for those specific regions.

It's important to note that the choice of piecewise transformations should be guided by an understanding of the data and the underlying relationships between variables. Segmenting the data and selecting appropriate transformations require domain knowledge and careful exploration of the dataset.



Correlation Causation and Confounding variables:

Let's say researchers want to study the impact of a new teaching method (independent variable) on students' test scores (dependent variable). However, the students' socioeconomic status (confounding variable) is not controlled for in the study. If students from higher socioeconomic backgrounds tend to perform better on tests, the confounding variable (socioeconomic status) may create a false impression that the new teaching method is more effective than it truly is.
Key Points:

- Confounding variables can distort the interpretation of study results.
- To address confounding, researchers may use study designs like randomized controlled trials or statistical methods like matching or regression analysis.
- Identifying and controlling for confounding variables is crucial for drawing accurate conclusions from research studies.

In summary, confounding variables are external factors that, if not properly accounted for, can lead to incorrect conclusions about the relationship between the variables being studied.

Forecasting with correlated predictors:

Having correlated predictors is not really a problem for forecasting, as we can still compute forecasts without needing to separate out the effects of the predictors. However, it becomes a problem with scenario forecasting as the scenarios should take account of the relationships between predictors. It is also a problem if some historical analysis of the contributions of various predictors is required.
Multicollinearity
Fortunately, if your purpose is primarily to predict or forecast Y, strong multicollinearity may not be a problem because a careful multiple regression program can still produce the best (least-squares) forecasts of Y based on all of the X variables.

From <https://www.sciencedirect.com/topics/mathematics/multicollinearity-problem>

Forecasts will be unreliable if the values of the future predictors are outside the range of the historical values of the predictors. For example, suppose you have fitted a regression model with predictors x_1 and x_2 which are highly correlated with each other, and suppose that the values of x_1 in the training data ranged between 0 and 100. Then forecasts based on $x_1>100$ or $x_1<0$ will be unreliable.
It is always a little dangerous when future values of the predictors lie much outside the historical range, but it is especially problematic when multicollinearity is present.
Note that if you are using good statistical software, if you are not interested in the specific contributions of each predictor, and if the future values of your predictor variables are within their historical ranges, there is nothing to worry about — multicollinearity is not a problem except when there is perfect correlation.

From <https://otexts.com/fpp3/causality.html>

Exponential smoothing

Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight. This framework generates reliable forecasts quickly and for a wide range of time series, which is a great advantage and of major importance to applications in industry.

From <https://otexts.com/fpp3/expsmooth.html>

a) **Simple exponential Smoothing**
   - When to be used :
   - No clear trend or seasonal patterns

It falls between naive an average method, one giving only weight to the last one and another giving average weights to all the observation.

Exponential smoothing falls between these two models,

- This is exactly the concept behind simple exponential smoothing. Forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past — the smallest weights are associated with the oldest observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots, \qquad (8.1$$

- where $0 \leq \alpha \leq 1$ is the smoothing parameter. The one-step-ahead forecast for time $T+1$ is a weighted average of all of the observations in the series $y_1, \ldots, y_T$. The rate at which the weights decrease is controlled by the parameter α.

For any alpha between 0 and 1 the weights attached to the observation decrease exponentially as we go back in time and the hence the name exponential smoothing.

Component form
where there is forecasting equation and smoothing equation.

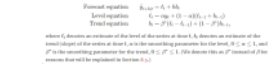
Now how to estimate alpha or the smoothing parameters.?
Choosing an appropriate value for αinvolves a trade-off between responsiveness to recent changes and stability based on past observations. There are a few common methods for determining the value of α:

1. Manual Tuning:
   - The choice of $\alpha$is often based on domain knowledge and the characteristics of the time series data.
   - Smaller values of α(e.g., 0.1) result in smoother forecasts that react more slowly to changes.
   - Larger values of α(e.g., 0.5 to 0.9) lead to more responsive forecasts that adapt quickly to recent changes.
2. Grid Search:
   - You can perform a grid search over a range of potential αvalues and evaluate their performance on a validation set or through cross-validation.
   - Iterate through different values of α(e.g., 0.1, 0.2, ..., 0.9) and select the one that provides the best forecasting accuracy.
3. Optimization Algorithms:
   - Some optimization algorithms can be used to automatically find the optimal value of αthat minimizes a forecasting error metric (e.g., Mean Squared Error) on a training dataset.
   - Techniques like gradient descent or other optimization algorithms can be employed to search for the best αvalue.

4. Exponential Smoothing Libraries:
   - Many time series forecasting libraries, such as statsmodels in Python, provide functions for automatic parameter selection. These functions may use optimization algorithms or heuristics to determine the optimal smoothing parameters.

From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

Holt's linear trend:
Holts extended the smooth exponential with a trend



Disadvantage of holts linear trend >
The forecasts generated by Holt's linear method display a constant trend (increasing or decreasing) indefinitely into the future. Empirical evidence indicates that these methods tend to over -forecast, especially for longer forecast horizons.

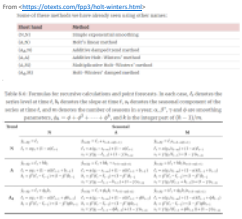
From <https://otexts.com/fpp3/holt.html>

So for this we introduce a damped parameter know as dampens
When $\phi = 1$ (damped parameter) then its equal to holts equation

In practice, ϕ is rarely less than 0.8 as the damping has a very strong effect for smaller values. Values of ϕ close to 1 will mean that a damped model is not able to be distinguished from a non-damped model. For these reasons, we usually restrict ϕ to a minimum of 0.8 and a maximum of 0.98.

From <https://otexts.com/fpp3/holt.html>
Holts _winters seasonal method consist of all three smoothing equation one for level and another one for trend and another for seasonal.
There are two variations to this method that differ in nature of the seasonal component
1) Additive method - when seasonal variation are constant throughout the seasons and
2) Multiplicative when the seasonal variations proportional to the level of the series.
With the additive method, the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component will add up to approximately zero. With the multiplicative method, the seasonal component is expressed in relative terms (percentages), and the series is seasonally adjusted by dividing through by the seasonal component. Within each year, the seasonal component will sum up to approximately m1

From <https://otexts.com/fpp3/holt-winters.html>



ETS(A,N,N): simple exponential smoothing with additive errors

From <https://otexts.com/fpp3/ets.html>
- estimate.
1. Error Correction Form:
- By rearranging the smoothing equation, we get the error correction form.
- It shows that the current level (ℓ) is adjusted based on the error in the previous forecast (er).
- The error (er) is the the difference between the actual observation ( ) and the previous forecast (-1).
- The adjustment is proportional to and the error (er). If the error is large, the adjustment is larger, and if the error is small, the adjustment is smaller.
- The adjustment corrects the estimated level based on how well the previous forecast aligned with the actual data.
In simpler terms, the model continuously learns and adjusts its estimate of the underlying level based on the errors in its past predictions. If the model consistently overestimates or underestimates, the adjustments become larger or smaller, respectively. The smoothing parameter ( ) controls the sensitivity to new observations, and a smaller leads to a smoother estimate.

The innovation state space model is a type of statistical model used for time series forecasting, and it belongs to the broader class of state space models. It shares some similarities with exponential smoothing models but also introduces additional components to capture various aspects of time series behavior. Here's a brief overview of both concepts:

1. **Exponential Smoothing Models:**
   - Exponential smoothing models, such as Simple Exponential Smoothing (SES), Double Exponential Smoothing (Holt's method), and Triple Exponential Smoothing (Holt-Winters method), are based on the idea of exponentially decreasing weights for past observations.
   - These models typically involve components like level, trend, and seasonality. Simple Exponential Smoothing considers only the level, Holt's method adds a trend, and Holt-Winters method incorporates both trend and seasonality.
   - The smoothing parameters control the weight given to recent observations, and the models make predictions based on weighted averages.

2. **Innovation State Space Model:**
   - The innovation state space model extends the state space modeling framework to capture innovations or shocks in the time series.
   - It involves two key components: the state equation and the observation equation.
   - **State Equation:** Describes how the underlying state of the system evolves over time, including potential trend or other latent components.
   - **Observation Equation:** Relates the observed data to the underlying state, accounting for any innovations or shocks that might occur.
   - The model explicitly models the innovations, allowing it to capture sudden changes or unexpected events in the time series.

Differences:
   - While exponential smoothing models focus on exponentially weighted averages and may include trend and seasonality components, the innovation state space model provides a more general framework that explicitly models innovations.
   - The state space model is more flexible and can handle various types of underlying structures and dependencies.
   - The innovation state space model is often used when there are irregular or abrupt changes in the time series that might not be well-captured by traditional exponential smoothing models.

In summary, the innovation state space model introduces a broader framework that allows for a more explicit representation of innovations or shocks in the time series, providing greater flexibility in capturing complex patterns and sudden changes. The choice between these models depends on the characteristics of the specific time series being analyzed.

Innovation state space model as explained to 5 year old
Imagine you have a magical toy that moves around and does different things. This toy has two special parts:
2. Magic Blueprint (State Equation): This is like the secret plan or recipe for how the toy changes over time. It says if the toy is going up, down, or staying the same. So, if the toy is a car, the blueprint might say if it's speeding up, slowing down, or cruising at the same speed.
3. Watching the Toy (Observation Equation): This part is like having a friend who watches the toy and tells you what's happening. Sometimes, the friend might see the toy doing unexpected things, like suddenly zooming faster or slowing down out of the blue.
Now, let's imagine the toy sometimes does really surprising stuff—maybe it suddenly jumps or spins around. The innovation state space model is like having a special way to understand and explain these surprises. It helps us figure out why the toy did something unexpected and understand the plan behind it.
In simpler terms, it's like having a cool way to understand and predict the surprises or changes in how the toy moves around. It's like having a magical guide that helps us make sense of the toy's actions, even when it does things we didn't expect!
So, the innovation state space model is a helpful tool to understand and predict unexpected or surprising events in a way that even a 5-year-old can grasp.

From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

4. Innovation Residuals:
   - Definition: Innovation residuals, also known as innovations or shocks, represent the unanticipated components of the time series. They capture the difference between the observed values and the expected or predicted values from the model.
   - Significance: Innovation residuals are essential for capturing sudden changes or unexpected events in the time series. They provide insight into the unexplained variations that the model hasn't accounted for explicitly.
5. Regular Residuals:
   - Definition: Regular residuals, often referred to simply as residuals, are the differences between the observed values and the predicted values of the time series based on the chosen model.
   - Significance: Regular residuals are useful for assessing the overall goodness-of-fit of the model. They indicate how well the model captures the observed data, and examining the distribution of residuals helps identify patterns or systematic errors.

From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

Estimation and model selection

Innovation means shock or sudden changes in the time series data

This emphasizes that ETS models, through their statistical framework, explicitly account for errors, making it possible to generate prediction intervals. The concept of prediction intervals is crucial for understanding the uncertainty associated with forecasts, and ETS models provide a more structured way to handle this uncertainty compared to traditional exponential smoothing methods.

Some source for python time series data :
https://alpersinbalc.medium.com/time-series-libraries-for-python-4e41e9d03288

ARIMA MODELS
Stationary time Series : Time series whose statistical properties do not depend in time , that is they do not have trends or seasonality.
Some times it maybe confusing as cyclic time series without trends and seasonality can be called stationary data

In general, a stationary time series will have no predictable patterns in the long -term.

Differencing: This is a technique to make non stationary time series into stationary one because it simplifies the modeling process and makers it easier to identify patterns and trends.

1. Definition:
   - Differencing involves computing the difference between consecutive observations in a time series.
2. First Order Difference:

- The most common form of differencing is the first-order difference, denoted as Δ��−��=��−1Δyt=yt−yt−1.
- This operation calculates the change between each data point and its immediate predecessor.

3. Purpose:
   - The goal of differencing is to remove the trend or seasonality present in the original time series, making it stationary.
   - Stationarity implies that the statistical properties of the time series, such as mean and variance, do not change over time.

(Plot such ACF helps identify non stationary time series.)

For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non -stationary data decreases slowly. Also, for non-stationary data, the value of Γ1 is often large and positive.

Random Walk:
Random Walking: Random walking is a concept used to describe a process where an entity, such as a variable or a position, moves in a sequence of random steps. Each step is independent of the previous ones, and the direction or size of the step is determined by a random element.
In financial terms, a random walk is often used to describe the movement of stock prices. The idea is that, in an efficient market, future price changes are not predictable, and each price movement is like a random step. The unpredictable nature of these steps is akin to a person taking random steps without any clear pattern.
Here's a simple analogy:
Imagine standing at a point and taking steps forward or backward, where the decision to move forward or backward is based on a flip of a coin. If it's heads, you take a step forward; if it's tails, you take a step backward. The sequence of steps you take forms a random walk.
Key Points:

- Independence: Each step is not influenced by the previous steps; it's independent.
- Unpredictability: The overall path becomes unpredictable over time, even though each step is determined randomly.
- Used in Finance: Random walks are often used as a theoretical model for the unpredictable nature of financial markets.

In financial theory, the idea of a random walk supports the efficient market hypothesis, suggesting that it's challenging to consistently outperform the market by predicting future price movements, as they are unpredictable and follow a random pattern.

From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

- A random walk is often considered a non-stationary time series because its mean and variance change over time, and it doesn't exhibit consistent patterns.
- Conversely, a stationary time series doesn't necessarily have to follow a random walk. It simply means that statistical properties remain constant, making it more amenable to modeling and forecasting.

In summary, while a random walk is a type of non-stationary time series, not all stationary time series behave like a random walk. Stationary series can have various patterns, trends, or seasonality without the unpredictable nature associated with a strict random walk.

From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

Financial data often exhibits a random walk or non -stationary behavior. To make it stationary, differencing is a common technique. Here's a breakdown:

4. First Differencing:
   - If the financial data follows a random walk, taking the first difference (subtracting each value from its previous one) can help remove the trend.
   - First differencing is often sufficient for making the series stationary if the underlying trend is linear.
5. Second Differencing:
   - In some cases, especially if the first difference still doesn't yield stationarity, you might need to take a second difference.
   - Second differencing involves differencing the already differenced series, providing a more pronounced effect on removing trends.

Remember that the goal is to achieve stationarity, where statistical properties remain constant over time. This makes it easier to apply time series models and draw meaningful insights. However, it's important to be cautious with excessive differencing. If you find yourself needing a high order of differencing, it might indicate that the data is highly unpredictable, and modeling challenges could persist. Always assess the results and choose the differencing order that best suits your specific dataset.
From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

There are Seasonal difference which can be done when there is strong seasonality in the data

But too much differentiation is also not good

Now how we check whether we need a differencing or not to make it stationary and for that we use
Unit root test
In this test, the null hypothesis is that the data are stationary, and we look for evidence that the null hypothesis is false. Consequently, small p -values (e.g., less than 0.05) suggest that differencing is required. From <https://otexts.com/fpp3/stationarity.html>

The KPSS test p-value is reported as a number between 0.01 and 0.1. If the actual p -value is less than 0.01, it is reported as 0.01; and if the actual p -value is greater than 0.1, it is reported as 0.1. In this case, the p-value is shown as 0.01 (and therefore it may be smaller than that), indicating that the null hypothesis is rejected. That is, the data are not stationary. We can difference the data, and apply the test again.From <https://otexts.com/fpp3/stationarity.html>

A similar feature for determining whether seasonal differencing is required is unitroot_nsdiffs(), which uses the measure of seasonal strength introduced in Section 4.3 to determine the appropriate number of seasonal differences required. No seasonal differences are suggested if Fs<0.64⁢⁢<0.64, otherwise one seasonal difference is suggested.

From <https://otexts.com/fpp3/stationarity.html>
Bachshift Operator or notation:
With back shift operator we are differing we can write like this :

The backward shift operator B⁢is a useful notational device when working with time series lags:Byt=yt−1.

In general, a d⁢th-order difference can be written as(1−B)dyt.

For example, a seasonal difference followed by a first difference can be written as(1−B)(1−Bm)yt

From <https://otexts.com/fpp3/backshift.html>

Auto regression Models

Autoregression, often denoted as AR, is a statistical modeling technique used in time series analysis. In simple terms, autoregression refers to modeling the relationship between a variable and its own past values. In other words, the value of the variable at the current time step is modeled as a linear combination of its previous values.
For example, an autoregressive model of order 1, denoted as AR(1), expresses the current value (⁢⁢) as a function of the previous value (⁢⁢−1Yt−1):
⁢⁢=⁢⁢+⁢⁢1⁢⁢−1+⁢⁢Yt=ϕ1Yt−1+εt
Here:

- ⁢⁢Yt is the value at time ⁢⁢.
- ⁢⁢1ϕ1 is a coefficient representing the impact of the previous value on the current value,
- ⁢⁢−1Yt−1 is the value at the previous time step,
- ⁢⁢εt is the error term, representing the random noise or unobserved factors affecting ⁢⁢Yt.

The autoregressive process continues for higher orders (AR(2), AR(3), and so on), involving more past values to predict the current value. Autoregressive models are widely used in time series forecasting to capture the temporal dependencies and patterns present in the data.
From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>
Moving Average Models >
A moving average model, often denoted as MA, is a statistical modeling technique used in time series analysis to capture patterns or dependencies in the data. Unlike autoregressive models that consider the relationship with past values of the variable itself, moving average models focus on the relationship between the current value and the past error terms.
The basic idea of a moving average model of order ⁢⁢q, denoted as MA(q), is to express the current value (⁢⁢Yt) as a weighted sum of the past ⁢⁢error terms (⁢⁢−1,⁢⁢−2,...,⁢⁢−1,⁢⁢−2,...,⁢⁢−q):
⁢⁢=⁢⁢+⁢⁢1⁢⁢−1+⁢⁢2⁢⁢−2+...+⁢⁢⁢⁢Yt=μ+εt+θ1εt−1+θ2εt−2+...+θqεt−q
Here:

- ⁢⁢Yt is the value at time ⁢⁢,
- ⁢⁢μ is the mean of the time series,
- ⁢⁢εt is the current error term, representing the random noise or unobserved factors affecting ⁢⁢Yt,
- ⁢⁢1,⁢⁢2,...,⁢⁢θ1,θ2,...,θq are coefficients representing the weights assigned to the past error terms.

The moving average process helps capture short-term fluctuations or random patterns in the time series. In combination with autoregressive models, moving average models contribute to creating more sophisticated models like ARMA (Autoregressive Moving Average) and ARIMA (Autoregressive Integrated Moving Average) for time series forecasting.

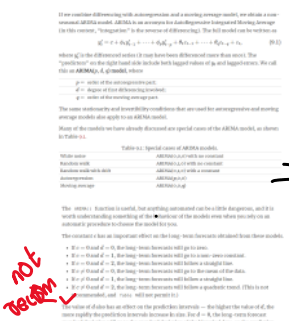
From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

Inertibility
imagine you have a sequence of numbers that represent some kind of data, let's say daily temperatures. A Moving Average (MA) model helps you understand how the current temperature depends on past random fluctuations.
Now, the term "invertible" is like saying you want to make this information more manageable and easy to understand. It's like taking a complex puzzle and turning it into a simpler one.
In the context of an MA model, being invertible means that you can look at the past random fluctuations in a way that makes sense. It's like saying, "Okay, these past influences don't have an endless impact on today's temperature." In other words, you can summarize the past in a neat way, making it easier to interpret and work with.
For example, if you find out that your MA model is not invertible, it's like saying, "Oops, the past influences are too chaotic, and I can't simplify them." It's a bit like having a messy puzzle that's hard to put together.
So, making an MA model invertible is like tidying up the information, making it more organized and easier to use for understanding and predicting future temperatures. It's about simplifying the complexity in a way that makes sense for analysis.

From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>

ARIMA

a) **Non-seasonal ARIMA models**



→ p, d, q (explained)

→ yt = εt (white noise)

→ yt = C + εt

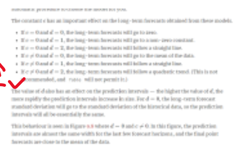
not needed

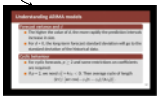
Difference between Moving Average and Moving Average Smoothing
Both moving average (MA) and moving average smoothing models are related concepts used in time series analysis, but they serv e different purposes.

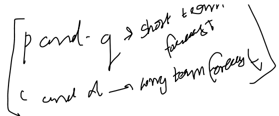
1. Moving Average (MA) Model:
   - Purpose: The MA model is a statistical model that describes the relationship between the current value of a time series and p ast error terms (residuals).
   - Equation: The general form of an MA model of order ⁢⁢is expressed as a weighted sum of past error terms.
   - Example Equation: ⁢⁢=⁢⁢+⁢⁢1⁢⁢−1+⁢⁢2⁢⁢−2+...+⁢⁢⁢⁢Yt=μ+εt+θ1εt−1+θ2εt−2+...+θqεt−q
   - Usage: It is used to capture short-term fluctuations or random patterns in the data.
2. Moving Average Smoothing:
   - Purpose: Moving average smoothing, on the other hand, is a technique used to reduce noise and highlight trends in a time seri es by computing the average of adjacent values.
   - Calculation: It involves taking the average of a specific number of consecutive observations (a window or period) to smooth o ut fluctuations.
   - Example Calculation: For a simple moving average of order ⁢⁢, each value in the smoothed series is the average of the current observation and the ⁢⁢−1⁢⁢−1previous observations.
   - Usage: It is used for trend analysis and noise reduction, making it easier to identify underlying patterns.

In summary, the MA model focuses on modeling the relationship between the current value and past error terms to capture short -term dependencies, while moving average smoothing is a data preprocessing technique that involves averaging neighboring values to create a smoother representation of the time series.

From <https://chat.openai.com/c/8d4a1495-0afe-469d-8276-c9b8d94e4feb>





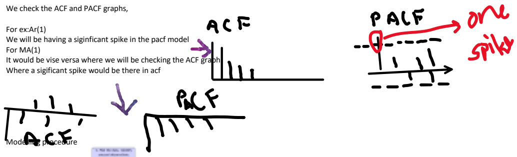


Partial Autocorrelation(PACF) and ACF graph

Now you have to select a pure AR model or pure MA model
i.e. ARIMA(p,0,0)and ARIMA(0,0,q)

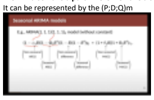
We check the ACF and PACF graphs,

For ex:Ar(1)
We will be having a siginficant spike in the pacf model
For MA(1)
It would be vise versa where we will be checking the ACF graph
Where a sigificant spike would be there in acf

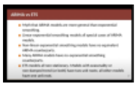


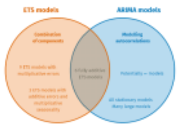
Modelling procedure

Seasonal ARIMA models
We will also have a part where we have a seasonal models.
It can be represented by the (P;D;Q)m

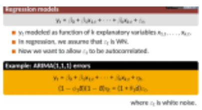


ARIMA VS ETS





DYNAMIC REGRESSION MODELS


Regression models

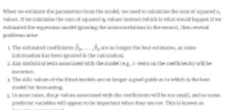
$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t,$$

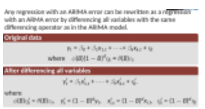
- y_t modeled as function of k explanatory variables $x_{1,t}, \ldots, x_{k,t}$.
- In regression, we assume that ε_t is WN.
- Now we want to allow ε_t to be autocorrelated.

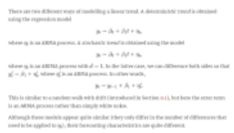
Example: ARIMA(1,1,1) errors

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \eta_t,$$
$$(1 - \phi_1 B)(1 - B)\eta_t = (1 + \theta_1 B)\varepsilon_t,$$

where ε_t is white noise.







More in the book

Hierarchical and groped time series :