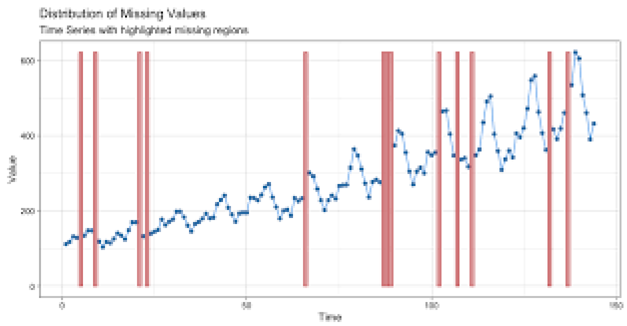
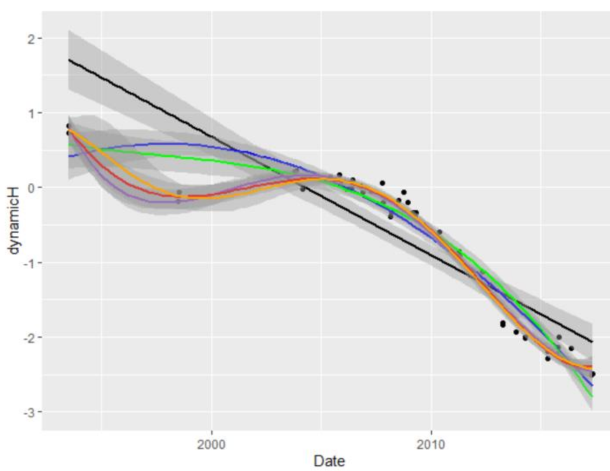
Lecture 21

Irregular Time Series
Dealing with irregular time series data can be challenging since it does not have a regular frequency, and observations may be missing or unevenly spaced. Here are some strategies for handling irregular time series:



Interpolation: interpolate missing values based on neighboring data points, such as linear interpolation or spline interpolation. This method can lead to a smoother time series, but it can also introduce biases. This method makes the most sense if the time series is mostly regular but with just a few missing values.



Resampling: resample the data into a regular frequency or a fixed interval, such as weekly or monthly, by aggregating or averaging the values within each interval. This method can simplify the data and reduce noise but may also result in information loss.

Time-aware models: use models that are specifically designed to handle irregular time series, such as point process models, which model events in continuous time, and state-space models, which can accommodate missing data points and irregular sampling intervals.

Feature engineering: create features that can capture temporal patterns in the data, such as lagged values, moving averages, and seasonality indicators.

Handling missing values: if the missing values are not too frequent, you can use methods like forward fill, backward fill, or interpolation to fill in the missing values. However, if the missing values are too frequent, you may need to consider models that can handle missing data, such as state-space models.

Outlier detection: irregular time series are more susceptible to outliers, so it is important to identify and remove them or adjust the model to accommodate them. This can be done by visual inspection, statistical tests, or using methods like robust regression or robust state-space models.

Overall, there is no one-size-fits-all solution for dealing with irregular time series, and the choice of method will depend on the specific characteristics of the data and the research question.

There are several potential problems associated with irregular time series analysis:

Missing data: Irregular time series may have missing data points, which can make it difficult to analyze trends and patterns in the data. Imputation or interpolation may be necessary to fill in missing values, but these methods may introduce biases and uncertainties.

Non-stationarity: Irregular time series may exhibit non-stationarity, where the mean and variance of the data change over time. This can make it difficult to model the data and make accurate predictions.

Outliers: Irregular time series may have outliers, which are data points that are significantly different from the rest of the data. Outliers can skew the results of the analysis and make it difficult to identify trends and patterns.

Measurement errors: Irregular time series may be subject to measurement errors, which can introduce noise and make it difficult to analyze the data. Careful data cleaning and preprocessing may be necessary to minimize the impact of measurement errors.

Data quality issues: Irregular time series may suffer from data quality issues, such as inconsistent formatting, missing metadata, or incorrect labeling. These issues can make it difficult to interpret the data and may require manual intervention to correct.

Sample size: Irregular time series may have a small sample size, which can limit the accuracy and generalizability of the analysis. Careful statistical methods and validation may be necessary to ensure the reliability of the results.

Interpolation and imputation techniques can be used to fill in missing values in a time series. These missing values can be due to various reasons, such as measurement errors, equipment failure, or missing data. Interpolation involves estimating the missing values based on the observed values of the time series at adjacent time points. This can be done using various methods, such as linear interpolation, cubic spline interpolation, or other interpolation methods.

Imputation, on the other hand, involves filling in the missing values based on other available information, such as the values of other variables that are correlated with the missing variable, or using statistical models to estimate the missing values. There are various imputation methods, such as mean imputation, regression imputation, multiple imputation, and Bayesian imputation.

However, it is important to note that interpolation and imputation can introduce bias and affect the accuracy of the time series analysis if not done carefully. Therefore, it is important to evaluate the validity of the imputed or interpolated values and assess the impact of missing data on the analysis results.

Imputing or interpolating missing values is an important step in time series analysis, as it allows for more complete data to be used in modeling and analysis. Some methods for imputing or interpolating missing values include:

Forward fill: This method involves filling missing values with the last observed value in the time series. This is a simple method that works well when the time series exhibits stable trends.

Backward fill: This method involves filling missing values with the next observed value in the time series. Like forward fill, this method works well when the time series exhibits stable trends.

Linear interpolation: This method involves fitting a line between the two nearest observed values and using that line to estimate missing values. This method works well when the time series exhibits linear trends.

Cubic spline interpolation: This method involves fitting a cubic polynomial curve between the two nearest observed values and using that curve to estimate missing values. This method works well when the time series exhibits non-linear trends.

Kalman filtering: This method uses a state-space model to estimate the missing values based on observed values and any available information about the underlying process generating the time series.

It is important to note that each method has its own strengths and weaknesses, and the choice of method will depend on the specific characteristics of the time series being analyzed. There are several diagnostics that can be used to assess model fit in a time series.

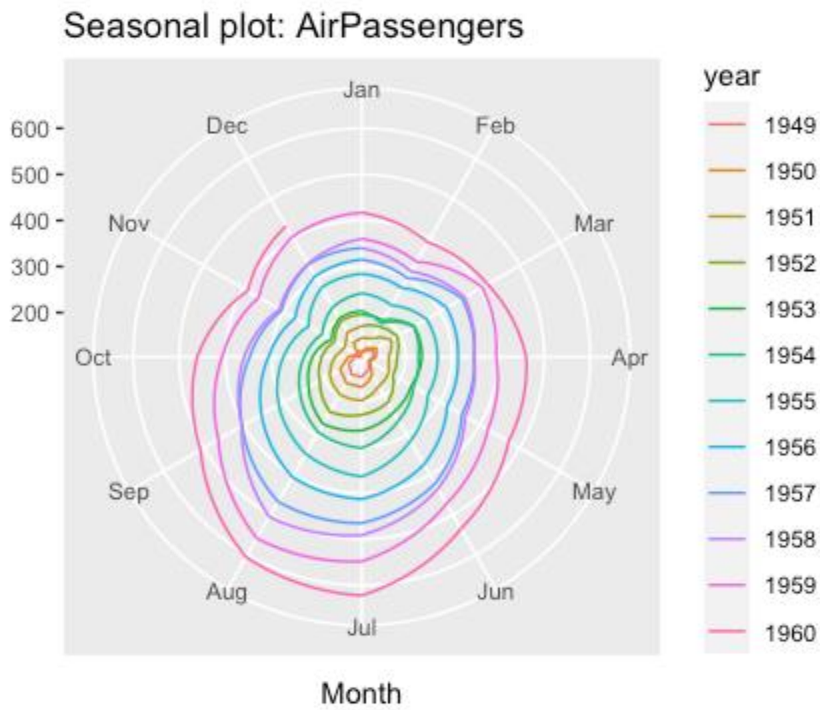
Residual plots: Plotting the residuals of the model against time can help identify patterns in the residuals that may suggest the presence of model misspecification or omitted variables. A good model should have residuals that are random and exhibit no patterns or trends. However, in time series models, errors can be correlated, which could show up on a residual plot.

Autocorrelation plots: Plotting the autocorrelation function (ACF) of the residuals can help identify any remaining correlation in the residuals. A good model should have residuals that are uncorrelated.
Ljung-Box test: This is a statistical test that measures the overall goodness-of-fit of a time series model by testing whether any autocorrelation in the residuals is statistically significant. A good model should have residuals that are not significantly autocorrelated.

Mean absolute error (MAE) and root mean squared error (RMSE): These are measures of the average difference between the predicted values and the actual values. A good model should have low values of MAE and RMSE.

Forecast accuracy: One of the most important diagnostics for time series models is forecast accuracy. This involves comparing the predicted values with the actual values for a holdout period. A good model should have accurate forecasts with low forecast errors.

Overall, it is important to use a combination of these diagnostics to assess model fit and ensure that the model is appropriate for the data being analyzed.



Next week, journal review

Resources:

1. http://math.furman.edu/~dcs/courses/math47/R/library/tseries/html/irts-methods.html
2. https://discuss.analyticsvidhya.com/t/time-series-forecasting-for-irregular-time-series-in-r/17574
3. https://www.statmethods.net/advstats/timeseries.html
4. https://subscription.packtpub.com/book/big-data-&-business-intelligence/9781783982028/12/ch12lvl1sec85/more-complex-time-series-objects
5. https://cran.r-project.org/web/packages/imputeTS/vignettes/imputeTS-Time-Series-Missing-Value-Imputation-in-R.pdf
6. https://medium.com/@poojamore282/an-introduction-to-missing-value-imputation-in-univariate-time-series-7739a34e87e3
7. https://www.mdpi.com/2073-4441/9/10/796
8. https://www.sciencedirect.com/science/article/abs/pii/S0925231221003003
9. https://pastas.readthedocs.io/en/v0.15.0/examples/017_Autocorrelation.ipynb.html
10. https://otexts.com/fpp2/residuals.html