GENERAL ASSEMBLY

17. TIME SERIES MODELING AND FORECASTING

PREVIOUSLY

- Time series
- Smoothing
- Trending and seasonality
- Stationarity, Dickey-Fuller test
- Autocorrelation, partial auto correlation
- Forecasting 101 & Metrics

TODAY

- ARMA
  Modeling
- ARMA
  Modeling

TYPES OF TS

Different types of TS

- White noise
- Trend
- Seasonality
- Cycle
- **Seasonality != Cycle**

Some cases can be confusing — a time series with cyclic behaviour (but not trend or seasonality) is NOT stationary. That is because the cycles are not of fixed length, so before we observe the series we cannot be sure where the peaks and troughs of the cycles will be.

STATIONARITY

Required for most models.

- Mean is constant $\mathrm{E}[Y_t] = \mu$
- Variance is constant $\mathrm{Var}(Y_t) = \mathrm{E}[(Y_t - \mu)^2] = \sigma^2$
- Autocorrelation is lag dependent

$$R(\tau) = \frac{\mathrm{E}[(Y_t - \mu)(Y_{t+\tau} - \mu)]}{\sigma^2}$$

TESTING FOR STATIONARITY

DICKEY-FULLER TEST

Null hypothesis: TS is NOT stationary

Demo in Notebook

- Dickey Fuller test does not test for seasonality stationarity

(PARTIAL) AUTOCORRELATION

ACF

Correlation between $Y_t|$ and $Y_{t-s}|$

PACF

Correlation between $Y_t|$ and $Y_{t-s}|$

- without the cumulative correlation between $Y_t|$ and $Y_{t-1} \cdots Y_{t-s+1}|$

SIMPLE FORECASTING

- next sample = last sample

$$\hat{Y}_{t+1} = Y_t$$

- Moving average

$$\hat{Y}_{t+1} = \frac{1}{n} \sum_{i=0}^{n} Y_{t-1-i}$$

- EWMA

$$\hat{Y}_t = \alpha \cdot Y_t + (1 - \alpha) \cdot \hat{Y}_{t-1}$$

- Linear Regression OLS

TODAY

- Transform TS into a Stationary TS
- Test is the TS predictable? Is it **white noise**?
- Decomposition: Trend, Seasonality, Residuals
- Is my forecast reliable?
- Is the Dow Jones a **Random Walk**?
- AutoRegressive modeling (AR) and Moving Average (MA)

DIFFERENCING

Create a new TS by taking the difference shifted by 1

$$X_t = Y_t - Y_{t-1}$$

! Try it out on the milk production ts

What happens to the seasonality? to the trend?

What is the result of the Dickey Fuller test on the difference? Is the difference series stationary?

WHITE NOISE

WHAT IS WHITE NOISE?

Time series data that shows **no auto correlation** is called **white noise**.

Formally, $X(t)$ is a white noise process if

- $E[X(t)] = 0$
- $E[X(t)^2] = \sigma^2$
- and $E[X(t)X(h)] = 0 \text{ for } t \neq h$

The autocorrelation matrix of a white noise TS is a diagonal matrix

HOW TO DETECT WHITE NOISE

1. ACF AND PACF

Rule of thumb:

- A Time series is white noise if 95% of the spikes in the Auto-correlation Function lie within $\pm \frac{2}{\sqrt{N}}$ with N the length of the time series.

=> Plot the PACF for the milk volume and difference TS and the tree rings series

Which one is a white noise?

TESTING FOR WHITE NOISE: THE LJUNG-BOX TEST

The Ljung–Box test may be defined as:

- H0: The data are independently distributed
- Ha: The data are not independently distributed; they exhibit serial correlation.

The test statistic is

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

where

- n is the sample size,
- $\hat{\rho}_k|$ is the sample autocorrelation at lag k,
- **h** is the number of lags being tested.

TESTING FOR WHITE NOISE: THE LJUNG-BOX TEST

Rule of thumb for h

- h = 10 for non-seasonal data
- h = 2m for seasonal data, where m is the period of seasonality.

RESIDUAL DIAGNOSTICS ON FORECASTING

A good forecasting method will yield residuals with the following properties:

- The residuals are uncorrelated: *If there are correlations between residuals, then there is information left in the residuals which should be used in computing forecasts.*
- The residuals have zero mean : *If the residuals have a mean other than zero, then the forecasts are biased.*

It is useful to also have the following two properties which make the calculation of prediction intervals easier

- The residuals have constant variance.
- The residuals are normally distributed.

These two properties make the calculation of **prediction intervals** easier

PREDICTION INTERVALS

95% prediction interval: $\hat{Y}_t \pm 1.96\sigma^2$ with σ an estimate of the standard deviation of the forecast distribution.

When the residuals are **normally distributed and uncorrelated** and when **forecasting one-step ahead**

=> the standard deviation of the *forecast distribution* is almost the same as the standard deviation of the *residuals*.

When conditions are not met, there are more complex ways to estimate confidence intervals

TS DECOMPOSITION

ADDITIVE MODEL

$$Y_t = S_t + T_t + E_t$$

where S_t is the seasonal component, S_t is the trend-cycle component and E_t is the residual

```
import statsmodels.api as sm
res = sm.tsa.seasonal_decompose(milk_prod.volume, model = 'additive')
resplot = res.plot()
```

MULTIPLICATIVE MODEL

$$Y_t = S_t \cdot T_t \cdot E_t$$

FORECAST WITH DECOMPOSITION

- forecast seasonality, trend and residuals separately
- add back together

LAB: IBM DATASET

Consider the daily closing IBM stock prices (data set ibmclose).

https://datamarket.com/data/set/2322/ibm-common-stock-closing-prices-daily-17th-may-1961-2nd-november-1962#!ds=2322&display=line

- Produce some plots of the data in order to become familiar with it.
- Split the data into a training set of 300 observations and a test set of 69 observations.
- Try various simple methods to forecast the training set and compare the results on the test set.
- Which method did best?

LAB: HOUSE SALES

https://datamarket.com/data/set/22q8/monthly-sales-of-new-one-family-houses-sold-in-th-e-usa-since-1973#!ds=22q8&display=line

Consider the sales of new one-family houses in the USA, Jan 1973 – Nov 1995 (data set hsales).

- Produce some plots of the data in order to become familiar with it.
- Split the data into a training set of 300 observations and a test set of 69 observations.
- Try various simple methods to forecast the training set and compare the results on the test set.
- Which method did best?

BREAK 5MN

NEXT SESSION

- R
- SQl
- AWS Machine Learning
- ?

LAB IS THE DJ A RANDOM WALK?

Why you cannot beat the market

WHAT'S A RANDOM WALK

When the differenced series is white noise, the model for the original series can be written as

$$y_t - y_{t-1} = e_t \quad or \quad y_t = y_{t-1} + e_t$$

A random walk model is very widely used for non-stationary data, particularly finance and economic data. Random walks typically have:

- long periods of apparent trends up or down
- sudden and unpredictable changes in direction.

http://python-for-signal-processing.blogspot.com/2014/04/random-walks-and-stumbles.html

NOTEBOOK: THE DOW JONES IS A RANDOM WALK

http://www.johnwittenauer.net/a-simple-time-series-analysis-of-the-sp-500-index/

- plot DJ
- plot diff
- transform with log
- plot rolling variance original + log
- plot diff of log => stationary time series model of daily changes to the S&P 500 index
- lag variables scatter plot => all centered and normal
- acf and pacf => no correlation => increment is white noise => we have a random walk
- decomposition of diff => look at the residuals white noise ?
- AR model, look at the residuals => much smaler values predicted than actual changes
- look at histogram of residuals
  - skewed => not great for confidence

- skewed => not great for confidence intervals

- autocorrelation plot of residuals
- test with Ljung-Box

AR(P) MODEL

In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable.

$$AR(1): \quad X_t = c + \varphi X_{t-1} + \varepsilon_t$$

$$AR(p): \quad X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

where

- φ_i are the parameters of the model
- ε_t is a white noise process with zero mean and constant variance σ_ε^2
- c is a constant

SPECIAL CASES

For an AR(1) model:

- When $\varphi_1 = 0$, X_t is equivalent to white noise.
- When $\varphi_1 = 1$, X_t is equivalent to a random walk.
- When $\varphi_1 = 0, c \neq 0$ X_t is equivalent to a random walk with drift
- When $\varphi_1 < 0$ X_t tends to oscillate between positive and negative values.

MA(Q) MODELS

Rather than use past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model.

Moving Average model of order q:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where

- μ is the mean of the series
- $\theta_1 \cdots \theta_q$ are the parameters of the model
- $\varepsilon_t \cdots \varepsilon_{t-q}$ are white noise error terms

ARIMA(P,D,Q) MODEL

We combine the AR(p) and the MA(q) and add i^th differencing

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

We call this an ARIMA(p,d,q) model, where

- p: order of the autoregressive part;
- d: degree of first differencing
  involved;
- q: order of the moving average part.

ESTIMATING P, D, Q

The *squirrel* approach

http://people.duke.edu/~rnau/411arim3.htm

The ML approach: Brute Force and Grid Search

CRITERIA FOR MODEL SELECTION

- Akaike Information Criterion (AIC)
- Schwarz Bayesian Information Criterion (BIC)
- Hannan-Quinn Information Criterion (HQIC)

WHITE NOISE - NORMALITY TEST

THE DURBIN WATSON TEST

The Durbin-Watson statistic ranges in value from 0 to 4.

- A value near 2 indicates non-autocorrelation;
- A value toward 0 indicates positive autocorrelation;
- A value toward 4 indicates negative autocorrelation.

AGOSTINO AND PEARSON FOR NORMALITY

Null hypothesis: the sample comes from a normal distribution

```
scipy.stats.normaltest(ts)
```

LAB ON SUNSPOTS

Wolf's Sunspot Numbers. 1700 – 1988

from https://bicorner.com/2015/11/16/time-series-analysis-using-ipython/

Time Series Classification and Clustering

DICKEY FULLER TEST

http://stats.stackexchange.com/questions/44647/which-dickey-fuller-test-should-i-apply-to-a-time-series-with-an-underlying-mode
http://stats.stackexchange.com/questions/225087/seasonal-data-deemed-stationary-by-adf-and-kpss-tests

RANDOM WALK

http://python-for-signal-processing.blogspot.com/2014/04/random-walks-and-stumbles.html http://fedc.wiwi.hu-berlin.de/xplore/tutorials/xegbohtmlnode39.html

LJUNG-BOX TEST

Thoughts on the Ljung-Box test
http://stats.stackexchange.com/questions/18135/testing-normality-and-independence-of-time-series-residuals

https://www.otexts.org/fpp/2/6

TRAIN TEST AND CROSS VALIDATION

https://www.otexts.org/fpp/2/5