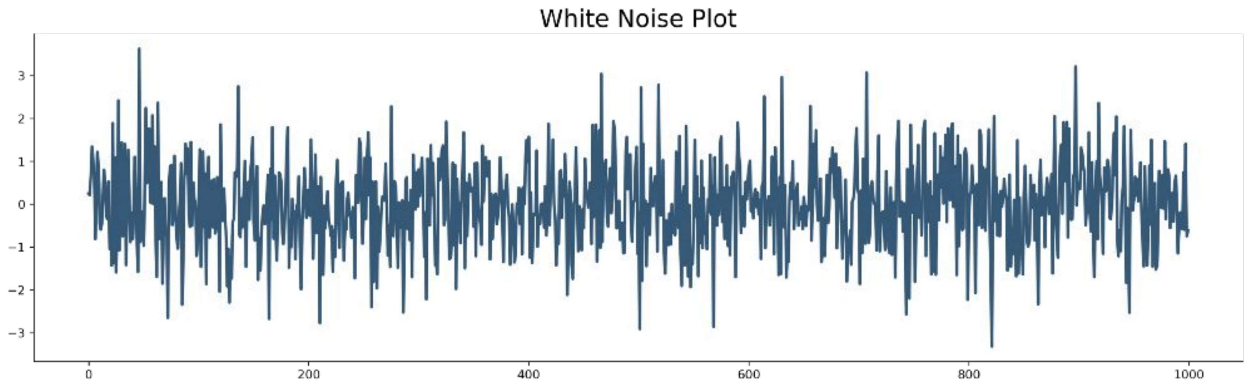
Auto Regression & Moving Average Forecasting

Stationary Timeseries

- Time series can be either stationary or nonstationary
- White noise is the strongest form of stationarity: zero mean, constant variance, and zero covariance between series values separated by a fixed lag.
- A strictly stationary: the probabilistic distribution of the values in any time interval [a, b] is identical to that in the shifted interval [a + h, b + h] for any value of the time shift h.



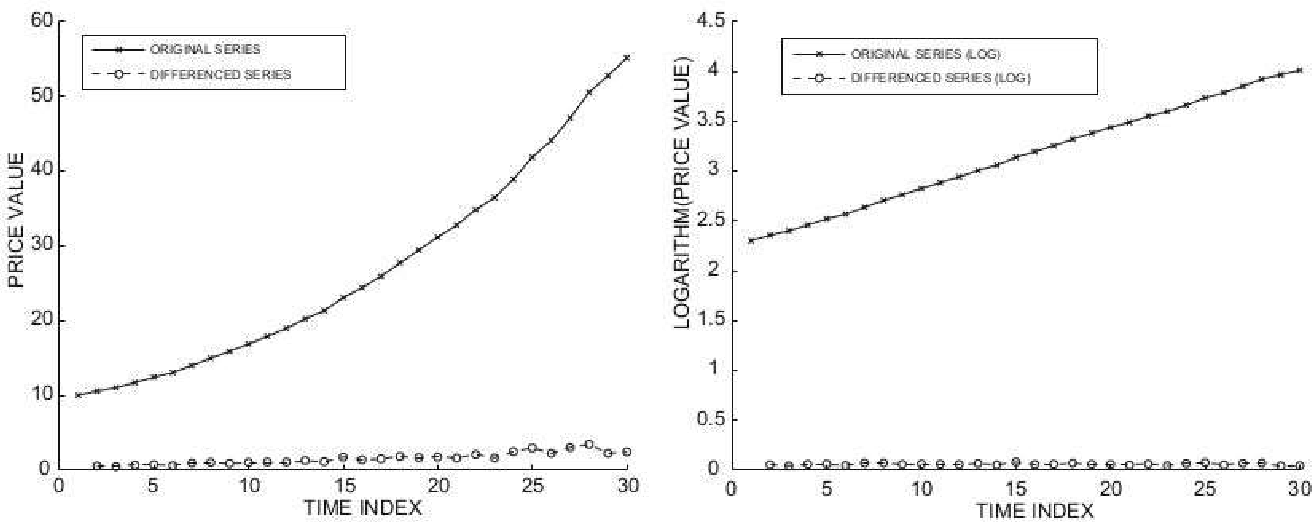
https://towardsdatascience.com/time-series-from-scratch-white-noise-and-random-walk-5c96270514d3

Stationary Timeseries

- In practice, timeseries are nonstationary.
- Forecasting can be better performed over stationary timeseries, since the statistical parameters (mean, variance, etc) do not vary over different time windows, and hence can be used in generating accurate forecasts.
- Therefore, it is often advantageous to convert nonstationary series to stationary ones before forecasting.
- After forecasting on the stationary series, the predicted values are transformed back to the original representation, using the inverse transformation.

Differencing - Converting Time Series to Stationary

- In differencing, the time series value y_i is replaced by the difference between it and the previous value. $y'_i = y_i - y_{i-1}$
- Differencing can be combined with log-scale



(a) Unscaled series (b) Logarithmic scaling

Differencing - Converting Time Series to Stationary

- Higher order differencing can be used to achieve stationarity in second order changes.
- Therefore, the higher order differenced value y_i is defined as follows:

$$y''_i = y'_i - y'_{i-1}$$
$$= y_i - 2 * y_{i-1} + y_{i-2}$$

- If the series is stationary after differencing, then an appropriate model for the data is:

$$y_{i+1} = y_i + c + e_{i+1}$$

- Here, e_{i+1} corresponds to white noise with zero mean, c is a non-zero constant that accounts for the drift
- It is rare to use differences beyond the second order.

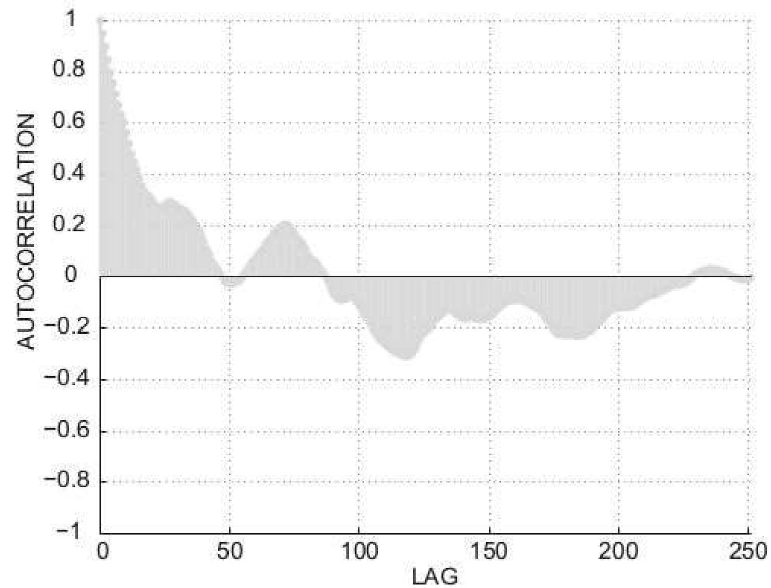
Autoregressive Models - Autocorrelation

- Use autocorrelation to predict future time series values (univariate forecasting)
- Autocorrelations represent the correlations between adjacently located timestamps in a series. Adjacency is defined in terms of *lag L*, e.g. Pearson coefficient,

$$\text{Autocorrelation}(L) = \frac{\text{Covariance}_t(y_t, y_{t+L})}{\text{Variance}_t(y_t)}$$

- Autocorrelation lies in the range [−1, 1]
- Typically positive for small values of L, and gradually drops, because nearby values in a timeseries tend to be similar.

Autocorrolation Function (ACF) plot



(a) IBM stock

Autoregressive Models AR(p)

- In the autoregressive model, the value of y_t at time t is defined as a linear combination of the values in the immediately preceding window of length p.

$$y_t = \sum_{i=1}^{p} a_i \cdot y_{t-i} + c + \epsilon_t$$

- A model that uses the preceding window of length p is referred to as an AR(p) model
- The parameters a_i, c are learned from the training data.

Autoregressive Models AR(p)

- In the autoregressive model, the value of y t at time t is defined as a linear combination of the values in the immediately preceding window of length p.

$$y_t = \sum_{i=1}^{p} a_i \cdot y_{t-i} + c + \epsilon_t$$

- A model that uses the preceding window of length p is referred to as an AR(p) model
- The parameters a_i, c are learned from the training data.
- The effectiveness of the forecasting model may be quantified by using the noise level in the estimated coefficients, for instance, it is desirable that the coefficient of determination R^2 is close to 1.

$$R^2 = 1 - \frac{\text{Mean}_t(\epsilon_t^2)}{\text{Variance}_t(y_t)}$$

Moving Average Models MA(q)

- The autoregressive models does not always explain all the variations, specially the unexpected component of the variations (shocks)
- This component can be captured with the use of a moving average model (MA)
- The autoregressive model can therefore be made more robust by combining it with an MA, so called ARMA model.
- The MA model is defined as $$y_t = \sum_{i=1}^{q} b_i \cdot \epsilon_{t-i} + c + \epsilon_t$$
- ε_{t-1} is a deviation from a predicted value, which can be viewed as white noise, or a shock
- AR relates the current value to the previous series values, while MA relates the current value to the previous history of deviations from forecasts.
- Series values cannot be predicted in terms of the shocks only. It needs to be added to the autocorrelations.

ARMA(p, q), ARIMA(p,d,q)

- The ARMA model is defined as $$y_t = \sum_{i=1}^{p} a_i \cdot y_{t-i} + \sum_{i=1}^{q} b_i \cdot \epsilon_{t-i} + c + \epsilon_t$$
- Recall that AR & MA can only work if the series is stationary
- For non-stationary series, we need to add differencing
- Combining differencing with ARMA leads to ARIMA *autoregressive integrated moving average model*
- If the order of the differencing is *d*, then this model is referred to as the ARIMA(p, d, q), e.g., when d=1

$$y'_t = \sum_{i=1}^{p} a_i \cdot y'_{t-i} + \sum_{i=1}^{q} b_i \cdot \epsilon_{t-i} + c + \epsilon_t$$

- ARIMA is thus the most general model for timeseries that can be made stationary.
- The predictors consist of lags of the dependent variable and/or lags of the forecast errors

Identifying p, d, q is Challenging

- More guidelines for your project here:

  https://people.duke.edu/~rnau/411arim2.htm

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

Forecasting at scale - the Prophet Model

- Recent model (and opensource) from Facebook, 2019
    - https://peerj.com/preprints/3190/
- The main design goals are: (1) ease of use, (2) speed
- At its core, the Prophet procedure is an additive regression model with four main components:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t.$$

  - g(t): a piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting changepoints from the data.
  - s(t): represents periodic changes (e.g., weekly and yearly seasonality), modeled using Fourier series.
  - h(t): represents the effects of holidays which occur on potentially irregular schedules over one or more days. The user provides the list of important holidays. Holiday list can be manually provided by user
  - ε_t: the error term represents the remainder which is not accommodated by the model, assumed to be normally distributed.

Credits & Readings

The slides of this lecture use material from

- CH14.3, Charu C.Aggarwal. Data Mining The Textbook, Springer.
- Robert Nau, Statistical forecasting: notes on regression and time series analysis

  https://people.duke.edu/~rnau/411arim.htm

- Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and Practice (3rd ed)

  https://otexts.com/fpp3/

- The prophet model: https://peerj.com/preprints/3190/