Time Series Models

While standard regression we must assume observations are independent from one another, but with time series data we expect that neighboring observations are correlated. Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. A time series is simply a set of statistics that is collected at regular intervals which we can use too obtain valid inferences. Ex. the daily number of live births or death.

A Single Observation of Stochastic Process

A **Stochastic process** is a (possibly) infinite sequence of variables ordered in time $\{Y_0, Y_1, Y_2 ...\}$. A time series is a single realization of a stochastic process. We want to make inference about the properties of the underlying stochastic process from a single observation.

There are two assumptions in time series analysis:

1. The data sequence is **stationary**. This means if all the times are shifted by the same amount, the probability distribution remains the same; meaning it depends on relative and not absolute values. In other terms:
   $(X_{t_1}, ... , X_{t_k}) =(X_{t_{1 + h}}, ... , X_{t_{k + h}})$
   for all time points *t* and integer *h*
   Under this assumption we can use the replication over time to make inferences about the common mean, variance, and other statistics. Additionally, the degree of independence increases as the time interval between two observations increases.
2. Ergodicity - the ability to make valid probability statements by looking over time rather than across replicates across one time.

Auto-correlation

A consequence of independence when observations are far apart enough in time is that we can use the *auto-correlation function* as a measure of dependence of the observations over time.

Let us define the covaraiance between time points at lag k as:
$\gamma(k) = Cov(Y_{k+1}, Y_1) = Cov(Y_{k+2}, Y_2)$

And from that we can define the autocorrelation function as:
$\rho(k) = \gamma(k) / \gamma(0) = corr(X_t, X_{t+k})$

We can also take a partial autocorrelation which is a conditional correlation. This is the correlation between two variables under the assumption that we know and take into account the values of some other set of variables.

Partial autocorrelation can be imagined as the correlation between the series and its lag, after excluding the contributions from the intermediate lags. So, PACF sort of conveys the pure correlation between a lag and the series. That way, you will know if that lag is needed in the AR term or not.

Auto-regressive Processes

An autoregressive process of order *p* is denoted by AR(*p*) is defined by

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} - \epsilon_t$$
with $\epsilon_t \sim N(0, \sigma^2)$

Or in other terms:
$$Y_t = \sum_{r=1}^{p} \phi_r Y_{t-r} + \epsilon_t$$
Where phis are a fixed constants

The outcome only depends on its own lags.

Moving Average Process

A time series Y_t is called a *moving average process of order q* (MA(q)) if:
$$Y_t = \sum_q^{s=0} \theta_s \epsilon_{t-s}$$
where thetas are a fixed constraint, theta_0 = 1

The sequence ϵ_t consisting of independent random variables with mean 0 and variance σ^2 are called **white noise.** It is a second order stationary series with $\gamma_0 = \sigma^2$ and $\gamma_k = 0$ for k != 0

The outcome depends only on the lagged forecast errors.

ARMA Models

ARIMA, short for ‘Auto Regressive Integrated Moving Average’ is actually a class of models that ‘explains’ a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

An ARIMA model is characterized by 3 terms: p, d, q

- p is the order of the AR term; It refers to the number of lags of Y to be used as predictors.
- q is the order of the MA term; the number of lagged forecast errors that should go into the ARIMA Model.
- d is the number of differencing required to make the time series stationary.

Revision #18
Created 6 April 2023 00:58:00 by Elkip
Updated 18 April 2023 16:32:51 by Elkip