**Time series**

***Stationarity:***

A time series is strictly stationary if the joint distribution of any collection of values remains the same regardless of the time at which they are observed.

P(Xt1​​=x1​, Xt2​​=x2 ​,…, Xtn​​=xn​) = P(Xt1​+k​=x1​, Xt2​+k​=x2​,…, Xtn​+k​=xn​), for any time shift k (the small x are fix values and yes they the same both sides)

In an easier conclusion of the above:

* The mean is constant
* The std/volatility is constant
* There is no seasonality (i.e., the covariance between two time periods depends only on the lag between them, not on the actual time at which the covariance is computed.) / autocorrelation does not change over time.
* No unit roots (ADF test of p-value<0.5)

**KPSS test**: Tests against the null hypothesis that the time series is stationary around a deterministic trend.

***Cointegration:***

Two or more non-stationary time series are combined in such a way that their linear combination is stationary.

**Engle-Granger Two-Step Method** for testing **cointegration** between two time series:

**1.Regress of one series on another**:

* Consider two non-stationary time series, Yt​ and Xt​.
* You perform a simple linear regression where one series is regressed on the other: Yt=α+βXt+ut.​ Here, α and β are coefficients, and ut​ is the residual of the regression.

The residual**s** ut​ represent the difference between the actual values of Yt​ and the values predicted by the linear relationship with Xt that is Yt hat.

If the ut, which is a time series in itself, is stationary, then ​Yt and Xt are cointegrated (i.e., have a long-term equilibrium relationship)

A black screen with colorful writing on it

Description automatically generated

**2. Estimate an Error Correction Model if cointegration is detected (to understand both the short-term impact and the long-term adjustment mechanism leading to the long-term relationship):**

Regress the first differences of Yt​ on the first differences of Xt​ and the lagged residuals (error correction term).

Here:

* = ​(first difference​)
* = ​(first difference​)
* ​ is the lagged residual from the long-run equation (error correction term).
* are coefficients representing the short-term relationship.
* is the error correction coefficient, which indicates the speed at which the series returns to equilibrium.
* is the error term of the ECM.

In short, if you want to predict Yt, you will use the ECM equation, not the first regression equation.