10.11.25

OPENCAMPUS.sh

Advanced Time Series Prediction

(S)ARIMA(X) + GARCH

- Organizational Matters
- Course Projects
- Topics:
  - (S)AR(I)MA(X)
  - (DCC)-GARCH

PROJECTS

- **Finance/Economics:**
  - …
- **Energy:**
  - ...
- **Environment:**
  - ...
- **Medical:**
  - ...
- **Engineering:**
  - ...

QUIZ: EXPLAIN THESE COMMANDS:

- df.head(), df.tail(), df.describe()
- df.isna(), df.isna().sum(), df.isna().sum(), df.asfreq()
- df.fillna(method = “ffill“) “bfill“
- df.fillna(value = df.mean())
- del df['dummy_variable']
- size int(len(df)*0.8)
- df_train= df.iloc[:size], df_test df.iloc[size:]
- df_train.tail()
- df_test.head()

QUIZ: EXPLAIN THESE VOCABULARY

- Missing values
- Seasonal Decomposition (Additive vs. Multiplicative),
- QQ-Plot, ACF, PACF
- Augmented-Dickey-Fuller Test, (Non)-Stationarity
- White Noise, Random Walk
- AR, MA, ARMA, ARIMA, ARIMAX, SARIMAX, GARCH

KEEP IN MIND:

- Weak-form stationary = covariance stationary
  - Cov(S_1)=Cov(S_2) i.e. WhiteNoise
  - Constant mean, Constant variance
  - Cov(x_n, x_n+k) = Cov(x_m, x_m+k)
- Strict Stationarity S_1 = S_2
  - Identical distributions, Rarely observed in nature/reality
- Stationary vs. Non-Stationary, Dickey-Fuller-Test, Augmented-Dickey-Fuller for time-dependent data, KPSS (more Power), Phililips-Perron,...

  - H_0: Phi_1 <1 Non-Stationarity is assumed in DF-Test
  - H_1: Phi_1=1
  - One-Lag Autocorrelation coefficient is lower than one.
  - Rejecting the null leads to stationarity
- sts.adfuller(df)

KEEP IN MIND:

- Go for parsimonious models via Log-Likelihood-Ratio-Test and AIC/BIC, lower is better
- Residuals ~ WhiteNoise (mu, sigma^2)
  - No patterns have been missed,
  - |Phi|or |Theta| > 1 otherwise factors blow up
- Forecasts from non-stationary data -> use transformations i.e. differencing
- including higher lags might lead to additional insignificant coefficients
  - LLR tests fails, or higher AIC/BIC values occur
- Normalization does not affect stationarity

FOR THE BRAVE:

SARIMAX (1, 0, 2) (2, 0, 1, 5)

$$y_t = c + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \Phi_1(y_{t-5} + \varphi_1 y_{t-6}) +$$

$$\Phi_2(y_{t-10} + \varphi_1 y_{t-11}) + \Theta_1(\varepsilon_{t-5} + \theta_1 \varepsilon_{t-6} + \theta_2 \varepsilon_{t-7}) + \varepsilon_t$$

$$AIC = -2\frac{\ln L}{T} + \frac{2}{T}k$$

$$BIC = -2\frac{\ln L}{T} + \frac{\ln(T)}{T}k$$

ln L → log likelihood of estimated model

k → # of parameters

T → length of time series

GARCH order (past σ_t^2)

↓

GARCH (1, 1)

↑

ARCH order (past ε_t^2)

MA order (past ε_t)

↓

ARMA (1, 1)

↑

AR order (past y_t)

ARMA - GARCH

μ_t = ARMA (p, q)

$\neq$

σ_t^2 = GARCH (p, q)

GARCH $(1, q)$ + ARMA $(3, q)$

FOR THE BRAVE:

Model-Specifications

AR-Model:

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

MA-Model:

$$y_t = c + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t$$

ARMA-GARCH-Model:

$$y_t = \mu + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t$$

$$\varepsilon_t = \sigma_t z_t \qquad \sigma_t^2 = \omega + \sum_{j=1}^{p'} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q'} \beta_j \sigma_{t-j}^2$$

FOR THE BRAVE:

GARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
EGARCH	$\log(\sigma_t^2) = \omega + \sum_{i=1}^{q} [\alpha_i e_{t-i} + \gamma_i(\lvert e_{t-i}\rvert - E\lvert e_{t-i}\rvert)] + \sum_{j=1}^{p} \beta_j \log(\sigma_{t-j}^2)$
GJR	$\sigma_t^2 = \omega + \sum_{i=1}^{q} [\alpha_i + \gamma_i I_{(\varepsilon_{t-i}>0)}]\varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$
IGARCH	$\sigma_t^2 = \omega + \varepsilon_{t-1}^2 + \sum_{i=2}^{q} (\varepsilon_{t-i}^2 - \varepsilon_{t-1}^2) + \sum_{j=1}^{p} \beta_j (\sigma_{t-j}^2 - \varepsilon_{t-1}^2)$
NGARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (\varepsilon_{t-i} + \gamma_i \sigma_{t-i})^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
TGARCH	$\sigma_t = \omega + \sum_{i=1}^{q} \alpha_i [(1-\gamma_i)\varepsilon_{t-i}^+ - (1+\gamma_i)\varepsilon_{t-i}^-] + \sum_{j=1}^{p} \beta_j \sigma_{t-j}$
APARCH	$\sigma^\delta = \omega + \sum_{i=1}^{q} \alpha_i [\lvert\varepsilon_{t-i}\rvert - \gamma_i \varepsilon_{t-i}]^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta$

REVIEW-QUIZ: EXPLAIN THESE QUESTIONS

- What are unit roots?
- What is the Lag-Operator and why is it useful?
- What can you say about the invertibility of time-series?
- What is the Augmented-Dickey-Fuller Test?
- What in Granger Causality?
- What is Cointegration?

1. **Unit Roots**: Unit roots are a characteristic of some time series, indicating that the series is non-stationary. A series with a unit root shows that its values persist over time and depend heavily on past values.
2. **Lag-Operator**: The Lag-Operator, denoted by L, is used in time series analysis to shift a time series backward by one or more periods. It's useful for simplifying the notation and algebra of time series models, particularly when modeling relationships at different time lags.
3. **Invertibility of Time-Series**: Invertibility refers to the condition under which a time series model can be rewritten or transformed in a way that expresses the current value as a function of past errors, rather than past values. This is crucial for simplifying the prediction and interpretation of the model.
4. **Augmented-Dickey-Fuller Test**: This test is a statistical method used to test if a time series has a unit root, which would suggest it is non-stationary. It's an extension of the Dickey-Fuller test that includes additional lagged difference terms to improve testing power.
5. **Granger Causality**: Granger causality is a statistical concept used to determine if one time series can predict another. If past values of one series (X) help predict future values of another series (Y) better than past values of Y alone, then X is said to Granger-cause Y.
6. **Cointegration**: Cointegration describes a scenario where two or more non-stationary series are linked in such a way that a linear combination of them is stationary. It often indicates a meaningful relationship where the series move together over time, despite each being non-stationary on their own.

1. **Unit Roots**:

$$(1 - L)^d y_t = \epsilon_t$$

Where y_t is the time series, L is the lag operator, d is the order of differencing, and ϵ_t is the error term. $d = 1$ in the case of a unit root.

2. **Lag-Operator**:

$$L^k y_t = y_{t-k}$$

Here, k is the number of periods the series is shifted back.

3. **Invertibility of Time-Series**:

$$y_t = \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}$$

In this equation, ψ_i are the weights given to past error terms ϵ. Invertibility depends on the condition that the ψ coefficients decay to zero as i increases.

4. **Augmented-Dickey-Fuller Test**:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \epsilon_t$$

This equation is used in the ADF test where Δy_t is the first difference of the series, α is a constant, β is the coefficient of a time trend, and ϕ_i are the coefficients of the lagged difference terms.

5. **Granger Causality**:

$$Y_t = \alpha + \sum_{i=1}^{n} \beta_i X_{t-i} + \sum_{i=1}^{n} \gamma_i Y_{t-i} + \epsilon_t$$

If the β_i coefficients are statistically significant, then X Granger-causes Y.

6. **Cointegration**:

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

If Y_t and X_t are non-stationary but the error term ϵ_t is stationary, Y_t and X_t are said to be cointegrated with cointegration vector $(1, -\beta)$.