

Time Series ARCH GARCH model

Business Objective

A time series is simply a series of data points ordered in time. In a time series, time is often the independent variable, and the goal is usually to make a forecast for the future.

Time series data can be helpful for many applications in day-to-day activities like:

- Tracking daily, hourly, or weekly weather data
- Monitoring changes in application performance
- Medical devices to visualize vitals in real-time

The ARCH (Autoregressive Conditional Heteroskedasticity) process introduced by Engle (1982) explicitly recognizes the difference between the unconditional and the conditional variance, allowing the latter to change over time due to past errors. Autoregressive: The current value can be expressed as a function of the previous values, i.e., correlated.

Conditional: This informs that the variance is based on past errors.

Heteroskedasticity: This implies the series displays unusual variance (varying variance).

As the name suggests, the GARCH is just the generalized version of the ARCH model. This generalization is expressed in including past variances as well as past squared residuals to estimate current (and subsequent) variances. The generalization comes from the fact that including a single past variance would (in theory) contain in itself the explanatory power of all other previous squared error terms. On the other hand, GARCH is a better fit for modelling time series data when the data exhibits heteroskedasticity but also volatility clustering. It serves as a sort of ARMA equivalent to the ARCH, where we're including both past values and past errors (albeit squared).

We have already covered the concepts of Autoregression modelling, Moving Average Smoothing techniques, ARIMA model, Multiple linear regression, and Gaussian process in the series till now.

In this project, we will be implementing the ARCH, GARCH models on the given dataset.

Data Description

The dataset is “Call-centres” data. This data is at month level wherein the calls are segregated at domain level as the call centre operates for various domains. There are also external regressors like no of channels and no of phone lines which essentially indicate the traffic prediction of the inhouse analyst and the resources available.

The total number of rows are 132 and number of columns are 8:

- Month, healthcare, telecom, banking, technology, insurance, no of phonelines and no of channels.

Aim

This project aims to build ARCH and GARCH models on the given dataset.

Tech stack

- Language - Python
- Libraries - pandas, numpy, matplotlib, seaborn, statsmodels, scipy, arch

Approach

1. Import the required libraries and read the dataset
2. Perform descriptive analysis
3. Data pre-processing
 - Setting date as Index
 - Setting frequency as month
4. Exploratory Data Analysis (EDA) -
 - Data Visualization
5. Perform train test split
6. Calculating returns and volatility
7. ARCH model
 - Install libraries
 - Build ARCH models with varying parameters
 - Build higher-lag ARCH models
8. GARCH model
 - Build a GARCH model
9. Forecasting the results
 - Forecast results on the best model

Modular code overview

```
input
|_CallCenterData.xlsx

src
|_Engine.py
|_ML_Pipeline
    |_ArchModel.py
    |_utils.py

lib
|_ARCH and GARCH.ipynb

output
|_Visualization plots(.png)
```

Once you unzip the modular_code.zip file, you can find the following folders within it.

1. input
2. src
3. output
4. lib
 1. Input folder - It contains all the data that we have for analysis. The following csv is used.
 - CallCenterData.xlsx
 2. Src folder - This is the most important folder of the project. This folder contains all the modularized code for all the above steps in a modularized manner. This folder consists of:
 - Engine.py
 - ML_PipelineThe ML_pipeline is a folder that contains all the functions put into different python files which are appropriately named. These python functions are then called inside the engine.py file.
 3. Output folder - The output folder contains all the visualization graphs. There are around 9 different plots.

4. Lib folder - This is a reference folder. It contains the original ipython notebook that we saw in the videos. The ppt used during the videos is also present here.

Project Takeaways

1. Introduction to Time series
2. Understand the basics of time series
3. Importing the dataset and required libraries
4. Data reprocessing
5. Exploratory Data Analysis (EDA)
6. What is a volatility measure?
7. Calculate returns and volatility.
8. What is the ARCH and GARCH models?
9. Building ARCH model
10. Building GARCH model
11. How to fit and train these models?
12. How to predict the model on test data?
13. How to interpret the results generated by these models?
14. Understand the mathematical notation of metrics
15. Configure ARCH and GARCH model using PACF plot