Projects - Financial econometrics

Topic 1. Test the effect of pandemic on the volatility in different countries
The aim is to investigate the effect of the COVID pandemic on the financial market in terms of volatility in different countries.
For the data set, you may choose the stock market indices from two countries of your choice for the period of 2017-2022.
You are expected to do a statistical test of the hypothesis of the equality of the volatility between two time series. Specifically, you will test (i) whether there is any difference in the volatility between 2017-2019 and 2020-2022 and (ii) whether there is any difference in the volatility between two countries in the same period.
You will be able to formulate the hypotheses in terms of the equality of the means of correlated data. It should be taken into account that the squared returns exhibit correlations, and thus the usual statistical test for equality between the means of two i.i.d. sequences should be modified accordingly. In particular, you will compute the p-value of the test result.

Topic 2. Estimation of a GARCH model & normality of the residuals
The aim is to study whether a GARCH model is a suitable model for returns data.
You are expected to develop codes to fit the GARCH (1,1) model, using the (pseudo) log-likelihood method,

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t, \quad (Z_t)_t \text{ i.i.d. } \mathcal{N}(0,1) \\ \sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \end{aligned}$$

For the data set, you may choose a financial time series of your choice and fit the model using the daily log returns.
You are expected to perform appropriate model diagnostics based on the residuals $(\hat{Z}_t)$ to evaluate the model assumptions of the normality of the conditional log returns.

Topic 3. Estimation of GARCH model & forecast of the volatility
The aim is to evaluate the performance of volatility forecasting by a GARCH model for financial time series data and see if there is any difference between different sectors of the industry.
You are expected to develop codes to fit the GARCH (1,1) model, using the (pseudo) log-likelihood method.

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t, \quad (Z_t)_t \text{ i.i.d. } \mathcal{N}(0,1) \\ \sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \end{aligned}$$

For the data set, you may choose financial time series (e.g. sector indices or any other) from two different sectors and fit the model using the daily log returns.

Using this model, predict the volatility of the series for the forthcoming days (stop the time horizon for the prediction when it is clear that the past data have no predictive power beyond that horizon). Plot the estimated daily volatility for the series during the pandemic year (2020-2022) and assess if there is any difference between two different sectors.

Topic 4. Estimation of a GARCH model & VaR computation

The aim is to compare and assess VaR estimates from the conditional and unconditional modelling approaches. Specifically, we consider the GARCH model as the conditional model and compare it with direct estimates from the filtered series (LT1). The latter assumes

$$\epsilon_t = \sigma_t Z_t, \quad (Z_t)_t \,\text{i.i.d}\, \mathcal{N}(0,1)$$

and the GARCH model assumes

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t, \quad (Z_t)_t \text{ i.i.d } \mathcal{N}(0,1) \\ \sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \end{aligned}$$

You are expected to develop codes to fit a GARCH model, using the (pseudo) log-likelihood method. For the data set, you may choose a financial time series of your choice and fit the model using the daily log returns for at least two years 2020-2021.

For each day of the year 2022, compute the 5% risk level VaR predicted from one day to the next one.

Assuming that we have used these predicted VaRs during the year 2022, count how many days the real loss of the asset would have been higher than the predicted VaR.

Topic 5. Estimation of T-GARCH model and study the leverage effect

The aim is to investigate whether an extension of GARCH-model is needed to account for asymmetric effect of the negative returns. The Threshold-GARCH (1,1) model is defined as

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t, \quad (Z_t)_t \text{ i.i.d. } \mathcal{N}(0,1) \\ \sigma_t^2 &= \omega + \alpha_-\epsilon_{t-1}^2 1_{\epsilon_{t-1}<0} + \alpha_+\epsilon_{t-1}^2 1_{\epsilon_{t-1}>0} + \beta\sigma_{t-1}^2. \end{aligned}$$

You are expected to develop codes to fit the T-GARCH model, using the (pseudo) log-likelihood method. And you will test the hypothesis :

$$H_0 : \alpha_- = \alpha_+$$

For the data set, you may choose a financial time series of your choice and fit the model using the daily log returns.

Topic 6. Intraday volatility estimation from high frequency data
The aim is to estimate the daily volatility from high frequency time series data and to assess the effect of the microstructure noise.
You are expected to estimate and plot the values of the estimated realized volatility when using observation frequencies ranging from 30 seconds to 15 mn. Compare these estimations with the long range estimation of the volatility (based on 1 month of daily data). Then, provide some estimation of the micro-structure noise size, by using the autocorrelation between the returns at different scales or any other appropriate methods.
Finding free high frequency data on the internet is rather difficult. It is possible to find tick data for the *Ishare S&P values (IVE)* on this web page : `http://www.kibot.com/free_historical_data.aspx`.
Tick data contains for each transaction, the time stamp and the price of the transaction. You may assume that the price of the asset at time t is given the value of the closest (in time) past transaction.
Finally, plot the evolution of the estimated daily volatility of the IVE over the last year.

Topic 7. Functional data analysis for intraday cumulative log returns
The aim is to understand and characterize the variability of the intraday cumulative log returns from high frequency time series data using functional data analysis techniques.
The data will be provided by the tutor, which contains log prices and time stamps (transformed into [0,1]) for several days of a high frequency time series.
Under the functional data framework, the data can be expressed as

$$Y_n(t_{nj}) = X_n(t_{nj}) + \varepsilon_{nj}\,, \quad n = 1, \ldots, N\,, j = 1, \ldots, J_n\,,$$

where $\varepsilon_{nj} \sim (0, \sigma^2)$ and

$$X_n(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{nk}\psi_k(t)\,.$$

You are expected to develop codes for performing functional PCA. With your estimates of $\hat{\mu}$, $\hat{\xi}_{nk}$ and $\hat{\psi}_k$, the fitted values are

$$\hat{X}_n(t) = \hat{\mu}(t) + \sum_{k=1}^{p} \hat{\xi}_{nk}\widehat{\psi}_k(t)\,.$$

Give an interpretation of the eigenfunctions and assess if $\hat{\xi}_{nk}$ follows a normal distribution for each k. Also examine if $(\hat{\xi}_{nk})_{n=1,\ldots,N}$ can be considered as a white noise process. Based on your analysis, give an estimate of the noise variance and assess the quality of the fit as a function of p. In addition, you may consider a time-series model for the sequence of the coefficients $(\hat{\xi}_{nk})_{n=1,\ldots,N}$ to develop a functional prediction method :

$$\widetilde{X}_{N+h|N}(t) = \hat{\mu}(t) + \sum_{k=1}^{p} \tilde{\xi}_{N+h,k}\widehat{\psi}_k(t)\,,$$

where $\tilde{\xi}_{N+h,k}$ is h-step ahead forecasts from $\hat{\xi}_{Nk}$. Evaluate the prediction performance in terms of mean square error. You may use the first few years of data to build your model and the later year to evaluate the prediction performance.