Homework

Backtesting

27.05.2020

Run the backtesting exercise using the historical data provided in three files (`Data01.dat`, `Data02.dat`, `Data03.dat`). Plot and inspect the distribution of normalized differences between the model prediction at **1 day horizon**. The model which prediction we would like to test is extremely simple - **Black-Sholes** model with **historically calibrated drift and volatility**. **The calibration window is 30 days wide**.

Backtesting is a powerful technique of testing the model prediction versus the historical realization. It can be considered as a final field test of the model.

As for the default parameters use the following:

Calibration window: 30 days,

Horizon: 1 day.

All other parameters should be chosen ad-hoc as needed.

As a good practice: make sure you leave all the parameters as inputs to the functions so you can modify them easily.

Make sure to illustrate all the points marked as *take notice* in the final project with the appropriate plots. Understand the final result.

Step by step guide:

1. To test the model predictions we first need to implement our model:

   (a) As Black-Sholes model predicts a relative changes while historical data consist of absolute numbers, hence we need to translate the historical data into set of relative movements. For each day (starting Day 2) calculate relative difference as :

   $$\text{RealizedRelativeDifference(t)} = \frac{\text{Value(t)} - \text{Value(t-1)}}{\text{Value(t-1)}}.[1]$$

   Take note of how plots of Date vs. Value and Date vs. RealizedRelativeDifference look like.

   (b) Starting with Day 32nd calibrate the model (relative data starts with Day 2):

[1] Checkout the *lag* function.

- Calculate drift as a slope parameter of linear regression (i.e. `RealizedRelativeDifference~Day`) over previous 30 days of data (historical window: Days 2-31).
- Estimate volatility as a standard deviation of a set of those 30 RealizedRelativeDifferences.

Take note how time series of those parameters looks like.

(c) Extrapolate the 30 days regression for Day 32 - it will be the expected value of our distribution with width equal to calibrated volatility. Bear in mind that model prediction is always a distribution (in this case Gaussian) and not a number.

(d) Regressing over days 2 - 31 gives a prediciton for day 32. Now slide 30 days window through the full data set (i.e. Day 33 prediction calculated by repeating the procedure on days 3 - 32, etc.).

2. From the expanded data set calculate the `Difference` and `NormalizedDifference` for each date (starting from 32nd).

$$\text{Difference(t)} = \text{Prediction(t)} - \text{RealizedRelativeDifference(t)},$$

$$\text{NormalizedDifference(t)} = \frac{\text{Difference(t))}}{\text{CalibratedVolatility(t)}}.$$

Take notice:

- How those differences evolve in time and what does that tells us.
- What does the values of differences looks when 'histogramed' and what does it tells us.
- Run the exercise on three data sets provided. What problems where you able to spot?

In case of questions do not hesitate to write me an email.