Volatility modeling and prediction

CentraleSupélec

June 2024

Table of contents

Outline of the project

- The aim of this project is to model and predict volatility of SPY financial time series.
- Part I will focus on computing volatility using different methods, mainly range-based volatility estimators to be able to predict it using classical time series models (ARCH, GARCH, etc).
- Part II will introduce Rough volatility models with their different calibration and prediction procedures.
- Part III will focus on establishing trading strategies using your volatility forecasts.

Evaluation

- Oral presentation + format of the presentation. Presentations are to be handed as pdf files.
- Code (commented notebook) must be handed in.
- Competences :
    - Group organization.
    - Added value.
    - Critical thinking.
    - Oral presentation

Table of contents

Data

- The SPDR SP 500 ETF Trust or SPY is one of the most popular funds that aims to track the SP 500 Index, which comprises 500 large-cap U.S. stocks.
- We will focus on SPY end of day Data from 2000 to end 2013.
- The data can be found on yahoo finance, which corresponds to opening, closing, high and low prices of each trading day.

Evolution of SPY closing prices

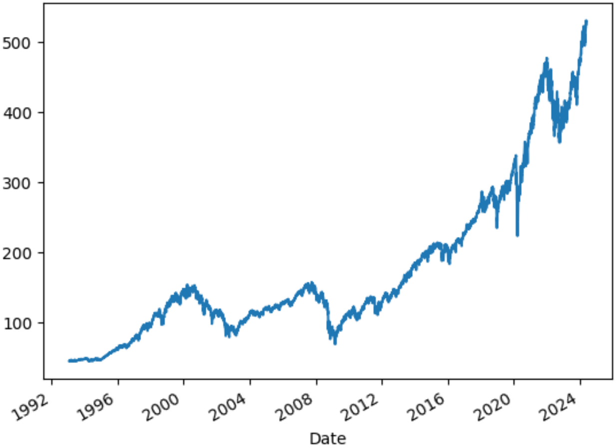

Figure – SPY closing prices from 1993 to 2024

Close-to-close volatility

- **Close-to-close volatility estimators** : Let $C_1, \ldots, C_T$ be the closing prices of an asset for T time periods $t = 1..T$[10].
  Then,
  $$\sigma_{cc,0}(T) = \sqrt{\frac{1}{T-1}\sum_{i=2}^{T} \ln \frac{C_i}{C_{i-1}}}$$
  is a biased [11] estimator of the asset volatility σ over the T time periods, assuming zero drift (i.e., $\mu = 0$), c.f. Parkinson.
  In addition,
  $$\sigma_{cc}(T) = \sqrt{\frac{1}{T-2}\sum_{i=2}^{T} \left(\ln \frac{C_i}{C_{i-1}} - \mu_{cc}\right)^2}$$
  with $\mu_{cc} = \frac{1}{T-1}\sum_{i=2}^{T} \ln \frac{C_i}{C_{i-1}}$, is a biased [11] estimator of the asset volatility σ over the T time periods, assuming non-zero drift (i.e., $\mu \neq 0$), c.f. Yang and Zhang.

Range-based volatility estimators

- Let be : $t = 1..T$, T time periods and $(O_1, H_1, L_1, C_1), \ldots, (O_T, H_T, L_T, C_T)$, the opening, highest, lowest and closing prices of an asset for time periods $t = 1..T$
- **Parkinson volatility estimator** Parkinson introduces an estimator for the diffusion coefficient of a Brownian motion without drift that relies on the highest and lowest observed values of this Brownian motion over a given time period.
  When applied to the estimation of an asset volatility, this gives the Parkinson volatility estimator $\sigma_P(T)$ defined over T time periods by

$$\sigma_P(T) = \sqrt{\frac{1}{T}} \sqrt{\frac{1}{4 \ln 2} \sum_{i=1}^{T} \left(\ln \frac{H_i}{L_i} \right)^2}$$

- **Garman-Klass volatility estimator** : Garman and Klass propose to improve the Parkinson estimator by taking into account the opening prices $O_t, t = 1..T$ and the closing prices $C_t, t = 1..T$.
  This leads to the Garman-Klass volatility estimator $\sigma_{GK}(T)$, defined over T time periods by

$$\sigma_{GK}(T) = \sqrt{\frac{1}{T}}\sqrt{\sum_{i=1}^{T}\frac{1}{2}\left(\ln\frac{H_i}{L_i}\right)^2 - (2\ln 2 - 1)\left(\ln\frac{C_i}{O_i}\right)^2}$$

Range-based volatility estimators

- **Rogers-Satchell volatility estimator** The Parkinson and the Garman-Klass estimators have both been derived under a zero drift assumption. When this assumption is not verified for an asset, for example because of a strong upward or downward trend in the asset prices or because of the usage of large time periods (monthly, yearly...), these estimators should in theory not be used because the quality of their volatility estimates is negatively impacted by the presence of a non-zero drift.
  In order to solve this problem, Rogers and Satchell devise the Rogers-Satchell volatility estimator $\sigma_{RS}(T)$, defined over T time periods by

$$\sigma_{RS}(T) = \sqrt{\frac{1}{T}} \sqrt{\sum_{i=1}^{T} \ln \frac{H_i}{C_i} \ln \frac{H_i}{O_i} - \ln \frac{L_i}{C_i} \ln \frac{L_i}{O_i}}$$

- **Yang-Zhang volatility estimator** The range-based volatility estimators discussed so far do not take into account opening jumps in an asset prices, that is, the potential difference between an asset opening price O_t and its closing price C_{t-1} for a time period t.
  This limitation causes a systematic underestimation of the true volatility. This leads them to introduce the multi-period [23] Yang-Zhang volatility estimator $\sigma_{YZ}(T)$, defined over T time periods by

$$\sigma_{YZ}(T) = \sqrt{\sigma_{ov}^2 + k\sigma_{oc}^2 + (1-k)\sigma_{RS}^2)}$$

where : $\sigma_{co}(T)$ is the close-to-open volatility, defined as

$$\sigma_{co} = \sqrt{\frac{1}{T-2}\sum_{i=2}^{T}\left(\ln\frac{O_i}{C_{i-1}} - \mu_{co}\right)^2}$$

with $\mu_{co} = \frac{1}{T-1}\sum_{i=2}^{T}\ln\frac{O_i}{C_{i-1}}$

Range-based volatility estimators

- σ_{oc} is the open-to-close volatility, defined as

$$\sigma_{oc}(T) = \sqrt{\frac{1}{T-2}\sum_{i=2}^{T}\left(\ln\frac{O_i}{C_i} - \mu_{oc}\right)^2}$$

  with $\mu_{oc} = \frac{1}{T-1}\sum_{i=2}^{T}\ln\frac{C_i}{O_i}$
- σ_{RS} is the Rogers-Satchell volatility estimator over the time periods $t = 2..T$
- k is chosen to minimize the variance the estimator.

Part I summary

- Estimate volatility using range-based estimators.
- Conduct a statistical study on the efficiency, precision and stability of each estimator.
- Predict intraday volatility using usual time series models and conduct a full statistical comparison study between models.

Table of contents

Main classes of volatility models

Prices are often modeled as continuous semi-martingales of the form

$$dP_t = P_t \left(\mu_t dt + \sigma_t dW_t \right).$$

The volatility process σ_s is the most important ingredient of the model. Practitioners consider essentially three classes of volatility models :

- Deterministic volatility (Black and Scholes 1973),
- Local volatility (Dupire 1994),
- Stochastic volatility (Hull and White 1987, Heston 1993, Hagan et al. 2002,...).

In term of regularity, in these models, the volatility is either very smooth or with a smoothness similar to that of a Brownian motion.

Rough Volatility model : geometric fOU

We make it formally stationary by considering a fractional Ornstein-Uhlenbeck model for the log-volatility denoted by X_t

$$dX_t = \nu dW_t^H + \alpha \left(m - X_t \right) dt$$

This process satisfies

$$X_t = \nu \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_t^H + m$$

We take the reversion time scale $1/\alpha$ very large compared to the observation time scale. This model is a particular case of the FSV model. However, in strong contrast to FSV, we take H small and $1/\alpha$ large. Thus we call our model Rough FSV (RFSV).

Intraday volatility estimation

We are interested in the dynamics of the (log)-volatility process. We use two proxies for the spot (squared) volatility of a day :

- A 5 minutes-sampling realized variance estimation taken over the whole trading day (8 hours).
- A one hour integrated variance estimator based on the model with uncertainty zones (Robert and R. 2012).

Note that we are not really considering a "spot" volatility but an "integrated" volatility. This might lead to some slight bias in our measurements (which can be controlled). From now on, we consider realized variance estimations on the S&P over 3500 days, but the results are fairly "universal".

Measure of the regularity of the log-volatility

The starting point of this work is to consider the scaling of the moments of the increments of the log-volatility. Thus we study the quantity

$$m(\Delta, q) = \mathbb{E}\left[\left|\log\left(\sigma_{t+\Delta}\right) - \log\left(\sigma_t\right)\right|^q\right],$$

or rather its empirical counterpart.

Let us first pretend that we have access to discrete observations of the volatility process, on a time grid with mesh Δ on $[0, T]$: $\sigma_0, \sigma_\Delta, \ldots, \sigma_{k\Delta}, \ldots$, $k \in \{0, \lfloor T/\Delta \rfloor\}$. Set $N = \lfloor T/\Delta \rfloor$, then for $q \geq 0$, we define

$$m(q, \Delta) = \frac{1}{N} \sum_{k=1}^{N} \left|\log\left(\sigma_{k\Delta}\right) - \log\left(\sigma_{(k-1)\Delta}\right)\right|^q$$

Distribution of log-volatility increments

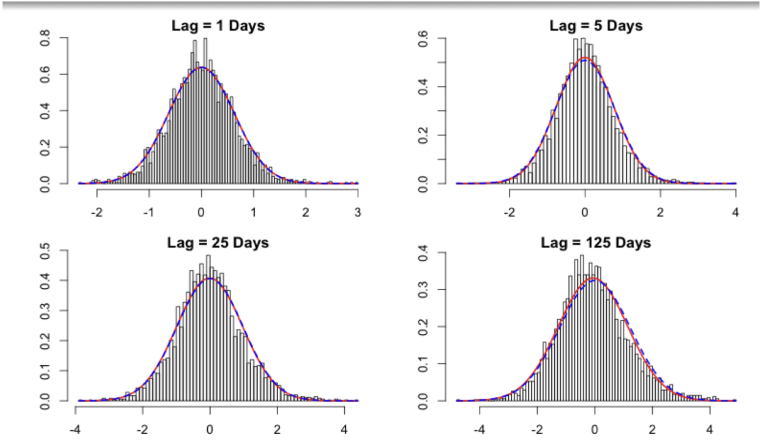

Figure – The distribution of the log-volatility increments is close to Gaussian.

Scaling of the moments

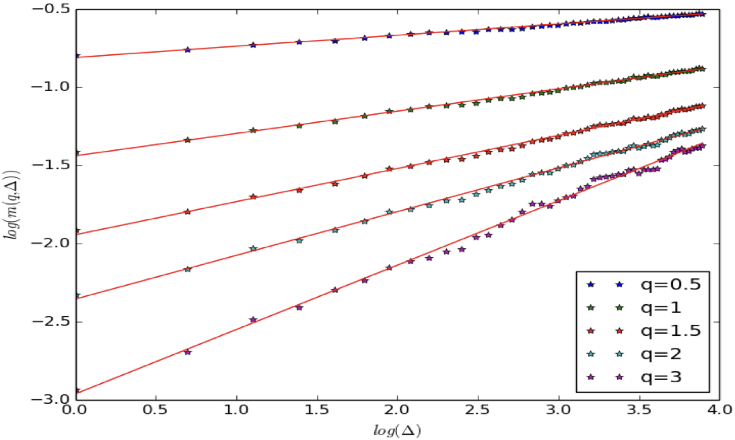

Figure – $\log(m(q,\Delta)) = \zeta_q \log(\Delta) + C_q$. The scaling is not only valid as Δ tends to zero, but holds on a wide range of time scales.

Monofractality of the log-volatility

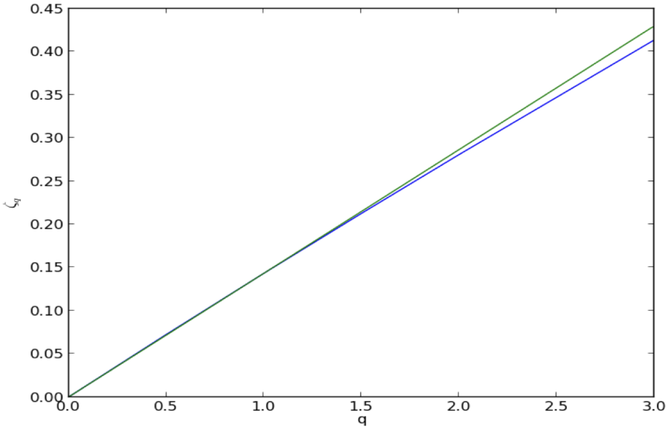

Figure – Empirical ζ_q and $q \to Hq$ with $H = 0.14$ (similar to a fBm with Hurst parameter H).

RFSV prediction formula

We apply the previous formula to the prediction of the log-volatility :

$$\mathbb{E}\left[\log \sigma^2_{t+\Delta} \mid \mathcal{F}_t\right] = \frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\int_{-\infty}^{t} \frac{\log \sigma^2_s}{(t-s+\Delta)(t-s)^{H+1/2}} ds$$

or more precisely its discrete version :

$$\mathbb{E}\left[\log \sigma^2_{t+\Delta} \mid \mathcal{F}_t\right] = \frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\sum_{k=0}^{N} \frac{\log \sigma^2_{t-k}}{(k+\Delta+1/2)(k+1/2)^{H+1/2}}$$

We compare it to usual predictors using the criterion

$$P = \frac{\sum_{k=1}^{N-\Delta}\left(\log\left(\widehat{\sigma^2_{k+\Delta}}\right) - \log\left(\sigma^2_{k+\Delta}\right)\right)^2}{\sum_{k=1}^{N-\Delta}\left(\log\left(\sigma^2_{k+\Delta}\right) - \mathbb{E}\left[\log\left(\sigma^2_{t+\Delta}\right)\right]\right)^2}$$

where $\mathbb{E}\left[\log\left(\sigma^2_{t+\Delta}\right)\right]$ denotes the empirical mean of the log-variance over the whole time period.

Part II summary

- Simulate a rough volatility process and conduct the different empirical tests, mainly the distribution of log-volatility increments, scaling of the moments and monofractality of the log-volatility.
- Same procedure for real market data.
- Study autocorrelation structure : autocorrelation of volatility ? log-autocorrelation of volatility ? autocorrelation of log-volatility ? With which scaling : Δ ? Δ^H ? Δ^{2H} ?
- Predict intraday volatility using Rough volatility modeling and conduct a statistical comparative study to other approaches.

Table of contents

How to find a trading strategy

- Consider a short/long position on an asset X and trading signal α. A base strategy is to buy asset X when $\alpha > 0$ and sell when $\alpha < 0$.
- Compute the returns of your investment and the cumulative P&L.
- Use metrics such as Sharpe Ratio and Drawdown.
- How about the risk of your strategy ? Can you quantify it ?
- You can use your data on SPY as well as your prediction of the volatility to trade VXX.
- Another strategy would be to simply trade SPY based on the volatility predictions. You can use historical data of SPY's returns and volatility, as well as other indices.