Computational Finance - Lab Assignment 1

Contact: k.chatziandreou@uva.nl // Submit until 18.04.2025

1 Financial Data Science: Realized Volatility [6p]

In this first assignment, we consider the problem of estimating the parameters μ and σ from market data under the stock price model (working under the physical measure $\mathbb{P}$):

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB_t. \tag{1}$$

Tasks: First, read AppendixA carefully. Then follow the steps:

1. **Step 1: Install Python Libraries to Retrieve Financial Data**
   In a command prompt or terminal, run:

```
pip install yfinance skfolio
```

   - `skfolio` is a Python library for portfolio optimization built on top of scikit-learn.
   - `yfinance` provides a way to fetch financial and market data from Yahoo! Finance.
   - We will primarily focus on very liquid tickers such as the SPX, though feel free to experiment with different ones.

2. **Step 2: Explore the yfinance API**
   Select a ticker of your choice and specify a start and end date (ideally covering a long historical period) to plot estimators for your mean and realized variance. For example:

```
import yfinance as yf
import datetime

ticker = "AAPL"
start_date = "2010-01-01"
end_date = datetime.datetime.now().strftime("%Y-%m-%d")

data = yf.download(ticker, start=start_date, end=end_date)
```

3. **Step 3: Historical Estimators**

   - Select a ticker and a start and end date (ideally a long time window) to plot estimators for the mean and realized variance using the adjusted closing price.
   - Compute the classical historical mean and volatility estimators (as referenced in equations (9) and (14) in your notes).
   - Implement one alternative volatility estimator of your choice (e.g., Parkinson or Garman–Klass) and compare it with the classical volatility estimator. For example, the Parkinson estimator is given by

$$\sigma_{\text{Parkinson}} = \sqrt{\frac{1}{4\ln 2}\sum_{t=1}^{T}\left[\ln\left(\frac{h_t}{l_t}\right)\right]^2}, \tag{2}$$

     where T is the number of days in the sample, h_t is the high price, and l_t is the low price on day t. Similarly, the Garman–Klass estimator is defined as

$$\sigma_{\text{Garman–Klass}} = \sqrt{\frac{1}{2T}\sum_{t=1}^{T}\left[\ln\left(\frac{h_t}{l_t}\right)\right]^2 - \frac{2\ln 2 - 1}{T}\sum_{t=1}^{T}\left[\ln\left(\frac{c_t}{o_t}\right)\right]^2}, \tag{3}$$

     where o_t is the open price and c_t is the close price on day t.
   - Compute a rolling window estimator using a 30-day window. Plot the results for all estimators and explain your findings.

- Create a volatility signature plot by displaying the time-series average of the realized variance

$$\bar{\sigma}^{(m)} = \frac{1}{T}\sum_{t=1}^{T} \hat{\sigma}_t^{(m)}$$

  as a function of the number of samples m (or equivalently, as a function of time in calendar or tick time). Plot your estimators as m varies and discuss the results.

- Finally, select one of the S&P 500 stocks and compare the realized variance estimators with the time series of implied volatilities for that ticker. To retrieve implied volatility data, you can use the `skfolio` package. For example:

```python
import pandas as pd
from skfolio.datasets import load_sp500_dataset, load_sp500_implied_vol_dataset
from skfolio.preprocessing import prices_to_returns

prices = load_sp500_dataset()
implied_vol = load_sp500_implied_vol_dataset()
X = prices_to_returns(prices)
X = X.loc["2010":]
implied_vol.tail()
```

4. **Step 4: Implied Volatility Data and VIX Estimation**

(a) Download the VIX quoted data for the specified time window:

```python
import yfinance as yf
import datetime

spx_symbol = "^SPX"
today = "2025-03-28"  # Keep this fixed in your implementation
end_date = today
start_date = end_date - datetime.timedelta(days=365)

spx_data = yf.download(spx_symbol, start=start_date, end=end_date)
lastBusDay = spx_data.index[-1]
vix_data = yf.download("^VIX", start=lastBusDay, end=lastBusDay
+ datetime.timedelta(days=1))
```

(b) To determine the available expiration dates for a particular ticker, call the `option_chain` attribute of the Ticker object. For example:

```python
import yfinance as yf

spx_ticker = yf.Ticker("^SPX")
# Suppose the next expiration is "2025-04-28"
expiry_date = "2025-04-28"  # Fixed to approximate a 30-day horizon as per CBOE
chain = spx_ticker.option_chain(expiry_date)
calls_df = chain.calls
puts_df  = chain.puts

print("Calls Head:")
print(calls_df.head())

print("Puts Head:")
print(puts_df.head())

# Optionally, save to CSV
calls_df.to_csv("spx_calls.csv", index=False)
puts_df.to_csv("spx_puts.csv", index=False)
```

(c) Compute the estimated VIX using the estimator VIX_t (referenced as equation (19) in your notes) and compare it with the CBOE-quoted VIX.

(d) Plot the historical estimated realized variances from **Step 3** alongside the VIX time series. Perform statistical analyses (such as correlation or cointegration tests) to assess the relationship between the time series.

(e) Run regression analyses between SPX returns and the VIX index, as well as between SPX returns and the historical realized variance estimator. Discuss your observations.
Hint: The variations of the stock index are typically negatively correlated with variations of the VIX index, whereas the correlation with historical volatility variations may differ.

2 Option Pricing: Power Option [4p]

Power options are useful in pricing realized variance and volatility swaps. Consider an asset whose price process $\{S_t\}_{t\geq 0}$ follows a geometric Brownian motion, which is the solution to the stochastic differential equation (SDE)

$$dS_t = \mu S_t \, dt + \sigma S_t \, dB_t,$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, and $\{B_t\}_{t\geq 0}$ is a standard Brownian motion.

Tasks:

1. **Solving the Black–Scholes PDE:**
Let $r \geq 0$ denote the risk-free interest rate. Solve the Black–Scholes partial differential equation (PDE)
$$r\, C(s,t) = \frac{\partial C}{\partial t}(s,t) + r\, s\, \frac{\partial C}{\partial s}(s,t) + \frac{\sigma^2}{2}\, s^2\, \frac{\partial^2 C}{\partial s^2}(s,t) \tag{4}$$
for $s > 0$ and $t \in [0, T]$, with the terminal condition
$$C(s,T) = s^2.$$
Hint: Try a solution of the form
$$C(s,t) = s^2 f(t),$$
and determine the function $f(t)$.

2. **Replicating Portfolio:**
Determine the portfolio holdings ξ_t and η_t in the risky asset S_t and the riskless asset A_t (with $A_t = A_0 e^{rt}$) such that the portfolio
$$V_t = C(S_t, t) = \xi_t S_t + \eta_t A_t, \quad 0 \leq t \leq T,$$
replicates the derivative contract with payoff
$$C(S_T, T) = (S_T)^2.$$

3 Hedging: Volatility Mismatch [8p]

Consider a short position in a European call option on a non-dividend paying stock with a maturity of one year and strike $K = 99$ EUR. Let the one-year risk-free interest rate be 6% and the current stock price be 100 EUR. Furthermore, assume that the volatility is 20%.
Use the Euler method to perform a hedging simulation.

Tasks:

1. **Matching Volatility:** Conduct an experiment where the volatility in the stock price process matches the volatility used in the delta computation (i.e., both set to 20%). Vary the frequency of hedge adjustments (from daily to weekly) and explain the results.

2. **Mismatched Volatility:** Perform numerical experiments where the volatility in the stock price process does not match the volatility used in the delta valuation. Run computational experiments for various levels of volatility and discuss the outcomes.

3. **Pricing and Hedging with Implied Volatility:** Assume that for an underlying with spot price S_t, a vanilla option with value function

$$C(t, S; \sigma_{\text{imp}}) := C_t,$$

is priced under the risk-neutral measure $\mathbb{Q}$ by solving the Black–Scholes–Merton PDE with implied volatility σ_{imp}:

$$rC_t - rS_t \frac{\partial C_t}{\partial S_t} - \frac{\partial C_t}{\partial t} - \frac{1}{2}\sigma_{\text{imp}}^2 S_t^2 \frac{\partial^2 C_t}{\partial S_t^2} = 0, \tag{5}$$

together with appropriate terminal conditions. The option delta is given by

$$\Delta_t = \frac{\partial C_t}{\partial S_t},$$

and is computed by solving (5). We also assume that the volatility σ_{imp} is implied from a market quote for $C(0, S; \sigma_{\text{imp}})$ and remains fixed until maturity T. This means the option is priced and delta-hedged at the level of the implied volatility, thereby avoiding exposure to directional changes in the underlying when hedging with a different volatility.

In practice, delta-hedging is performed in discrete time (typically daily), and real securities do not follow perfect log-normal diffusive processes. Thus, assume that the dynamics of the underlying price S_t under the objective (real-world) measure $\mathbb{P}$ are given by

$$\frac{dS_t}{S_t} = r\,dt + \sigma_t\,dW_t^{\mathbb{P}}, \tag{6}$$

where σ_t denotes the instantaneous realized volatility. Model parameters in this SDE are estimated for trading and risk management purposes, and the expected profit and loss (P&L) and its variance are computed under $\mathbb{P}$.

Show that in order for the final P&L to vanish on average, the implied volatility σ_{imp} used for pricing and risk management must, on average, match the future realized volatility σ_t when weighted by the option's dollar gamma over its life. That is, one must have

$$\mathbb{E}\left[\int_0^T e^{-rt}\frac{1}{2}S_t^2\,\frac{\partial^2 C_{\sigma_{\text{imp}}}}{\partial S_t^2}\left(\sigma_{\text{imp}}^2 - \sigma_t^2\right)dt\right] = 0, \tag{7}$$

where σ_t is the instantaneous realized volatility.

A Assignment 1

Historical Trend Estimation:
By discretizing (1) along a partition of the interval $[0, T]$ at observation dates $t_0, t_1, \ldots, t_N$, we obtain

$$\frac{S_{t_{k+1}} - S_{t_k}}{S_{t_k}} = \mu(t_{k+1} - t_k) + \sigma(B_{t_{k+1}} - B_{t_k}), \quad k = 0, 1, \ldots, N-1. \tag{8}$$

A natural estimator for the drift parameter μ is given by

$$\hat{\mu}_N := \frac{1}{N}\sum_{k=0}^{N-1} \frac{1}{t_{k+1} - t_k}\left(\frac{S^M_{t_{k+1}} - S^M_{t_k}}{S^M_{t_k}}\right), \tag{9}$$

where $(S^M_{t_{k+1}} - S^M_{t_k})/S^M_{t_k}$, $k = 0, 1, \ldots, N-1$, denotes market returns observed at discrete times $t_0, \ldots, t_N$.
Historical Log-Returns:
Alternatively, by considering log-returns, we have

$$\begin{aligned}
\log \frac{S_{t_{k+1}}}{S_{t_k}} &= \log S_{t_{k+1}} - \log S_{t_k} \\
&= \log\left(1 + \frac{S_{t_{k+1}} - S_{t_k}}{S_{t_k}}\right) \\
&\approx \frac{S_{t_{k+1}} - S_{t_k}}{S_{t_k}}.
\end{aligned}$$

With uniform time increments $t_{k+1} - t_k = \frac{T}{N}$, we can replace (9) with the telescoping estimator:

$$\frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{t_{k+1}-t_k}\left(\log S_{t_{k+1}}-\log S_{t_k}\right)=\frac{1}{T}\log\frac{S_T}{S_0}. \tag{10}$$

Using (9), the main contribution to the variance of $\hat{\mu}_N$ is

$$\operatorname{Var}(\hat{\mu}_N)=\frac{1}{T^2}\sum_{k=0}^{N-1}\operatorname{Var}(\sigma\Delta B_t)=\frac{\sigma^2}{T}, \tag{11}$$

which implies a standard deviation of $\sigma/\sqrt{T}$. To construct a 95% confidence interval with a 1% window (i.e., $\pm 0.5\%$), we set:

$$q(\alpha)\frac{\sigma}{\sqrt{T}}\leq 0.5\%, \tag{12}$$

leading to $T\geq(1.96\sigma/0.005)^2$. Thus, for a volatility $\sigma=0.2$ (20%), we require more than 6.146 years to obtain an unbiased drift estimator with a precision of 1%. This considerable timeframe motivates our primary focus on estimating realized volatility rather than drift.

Historical Volatility Estimation:

The volatility parameter σ can be estimated by rearranging (8):

$$\sigma^2\sum_{k=0}^{N-1}\frac{(B_{t_{k+1}}-B_{t_k})^2}{t_{k+1}-t_k}=\sum_{k=0}^{N-1}\frac{1}{t_{k+1}-t_k}\left(\frac{S_{t_{k+1}}-S_{t_k}}{S_{t_k}}-(t_{k+1}-t_k)\mu\right)^2, \tag{13}$$

leading to the unbiased realized volatility estimator:

$$\hat{\sigma}_N^2:=\frac{1}{N-1}\sum_{k=0}^{N-1}\frac{1}{t_{k+1}-t_k}\left(\frac{S_{t_{k+1}}-S_{t_k}}{S_{t_k}}-(t_{k+1}-t_k)\hat{\mu}_N\right)^2 \tag{14}$$

$$=\frac{1}{N-1}\sum_{k=0}^{N-1}\frac{1}{t_{k+1}-t_k}\left(\frac{S_{t_{k+1}}-S_{t_k}}{S_{t_k}}\right)^2-\frac{T}{N-1}\hat{\mu}_N^2. \tag{15}$$

The VIX Index:

An alternative approach to estimate market volatility is using the CBOE Volatility Index (VIX), particularly for the S&P 500 Index (SPX). Consider an asset price process S_t satisfying

$$dS_t=rS_tdt+\sigma_tS_tdB_t, \tag{16}$$

which can also be represented as

$$S_t=S_0\exp\left(\int_0^t\sigma_sdB_s+rt-\frac{1}{2}\int_0^t\sigma_s^2ds\right), \tag{17}$$

where $(\sigma_t)_{t\geq 0}$ denotes a stochastic volatility process.

Lemma A.1 *Let $\phi\in C^2((0,\infty))$. For all $y>0$, we have the Taylor-type representation*

$$\phi(x)=\phi(y)+(x-y)\phi'(y)+\int_0^y(z-x)^+\phi''(z)dz+\int_y^\infty(x-z)^+\phi''(z)dz \tag{18}$$

for all $x>0$.

The next proposition shows that the VIX volatility index, defined by

$$\mathrm{VIX}_t:=\sqrt{\frac{2e^{r\tau}}{\tau}\left(\int_0^{F_{t,t+\tau}}\frac{P(t,t+\tau,K)}{K^2}dK+\int_{F_{t,t+\tau}}^\infty\frac{C(t,t+\tau,K)}{K^2}dK\right)}, \tag{19}$$

at time $t>0$, can be interpreted as an average of future volatility. Here, according to CBOE documentation, $\tau=30$ days, and

$$F_{t,t+\tau}:=\mathbb{E}_{\mathbb{Q}}[S_{t+\tau}\mid\mathcal{F}_t]=e^{r\tau}S_t \tag{20}$$

represents the forward price at time $t+\tau$. $P(t,t+\tau,K)$ and $C(t,t+\tau,K)$ denote out-of-the-money put and call option prices at maturity $t+\tau$ with strike K. One can further show that the VIX index at $t\geq 0$ corresponds to the averaged realized variance swap price:

$$\mathrm{VIX}_t = \sqrt{\frac{1}{\tau}\mathbb{E}_{\mathbb{Q}}\left[\int_t^{t+\tau} \sigma_u^2 du \mid \mathcal{F}_t\right]}. \tag{21}$$

The second goal is to estimate the VIX index based on the discretization of (19) and market option prices on the SPX. The strikes for OTM puts and calls are ordered as follows:

$$K_1^{(p)} < \cdots < K_{n_p-1}^{(p)} < K_{n_p}^{(p)} := F_{t,t+\tau} =: K_0^{(c)} < K_1^{(c)} < \cdots < K_{n_c}^{(c)}, \tag{22}$$

and the discretization of (19) becomes:

$$\begin{aligned}
\mathrm{VIX}_t^2 &= \frac{2e^{r\tau}}{\tau}\left(\sum_{i=1}^{n_p-1}\int_{K_i^{(p)}}^{K_{i+1}^{(p)}} \frac{P(t,t+\tau,K_i^{(p)})}{K^2}dK + \sum_{i=1}^{n_c}\int_{K_{i-1}^{(c)}}^{K_i^{(c)}} \frac{C(t,t+\tau,K_i^{(c)})}{K^2}dK\right) \\
&= \frac{2e^{r\tau}}{\tau}\left(\sum_{i=1}^{n_p-1} P(t,t+\tau,K_i^{(p)})\left(\frac{1}{K_i^{(p)}} - \frac{1}{K_{i+1}^{(p)}}\right) + \sum_{i=1}^{n_c} C(t,t+\tau,K_i^{(c)})\left(\frac{1}{K_{i-1}^{(c)}} - \frac{1}{K_i^{(c)}}\right)\right).
\end{aligned}$$