Statistical Machine Learning I - 2025-2026
TP 3

Exercise 1

The ARCH(1) model for financial time series (Engle, 1982) is defined by

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t \quad \text{with} \quad \epsilon_t \sim \mathrm{N}(0,1), \\ \text{where } \sigma_t^2 &= \theta_1 + \theta_2 Y_{t-1}^2, \quad \theta_1 > 0,\ \theta_2 > 0. \end{aligned}$$

ARCH is the acronym for Auto Regressive Conditionally Heteroschedastic. Its goal is to model *volatility clustering* ("Conditionally Heteroschedastic": the conditional variance given the past is not constant): it is the phenomenon observed in finance that high volatility (high variance) tends to occur in cluster, as do quiet periods (low volatility) on the financial markets.

The ARCH model is used in finance to model log-returnx $Y_t = \log(X_{t+1}/X_t)$ of stock option values or indexes X_t such as the Nasdaq.

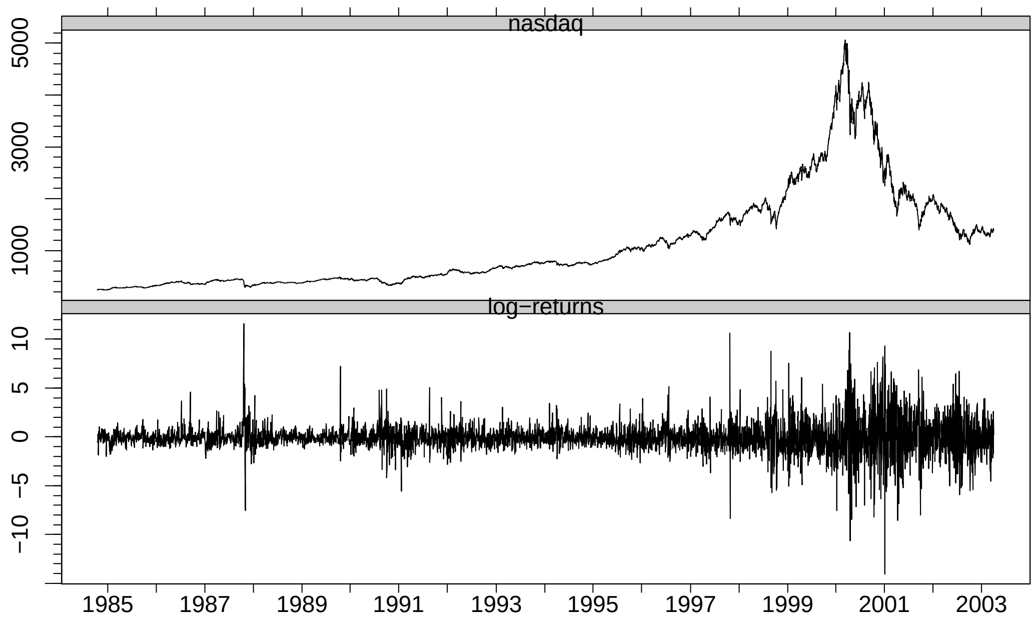

Figure 1: Time series of NASDAQ prices X_t (top) and its log-returns Y_t as a function of time.

1. Write a function `simuARCH1` that simulates data, here a time series of length n, that follows an ARCH(1) model of parameter (θ_1, θ_2).

2. Derive the likelihood function (conditional on Y_1) and write a function `neglogL-ARCH1` that calculates for any $(\theta_1, \theta_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ the negative log-likelihood.

3. Write a function `MLE-ARCH1` that calculates the MLE $(\hat{\theta}_1^{\mathrm{MLE}}, \hat{\theta}_2^{\mathrm{MLE}})$ given data.

4. Choose a set of value for (θ_1, θ_2), preferably with $\theta_2 < 1$ for the time series to be stationary. Perform a Monte-Carlo simulation to investigate the asymptotic normality of the of $\hat{\theta}_1^{\mathrm{MLE}}$ and $\hat{\theta}_2^{\mathrm{MLE}}$ predicted by the theory.

5. For $\theta_2 > 1$, the time series is known to not be stationary. For the sample sizes n that you considered in Point 4 above and with a true $\theta_2 = .8$, what proportion of $\hat{\theta}_2^{\text{MLE}}$ tend to show that the time series is not stationary?