M2 Quantitative Finance

Financial Econometrics

Estimating the Value-at-Risk, analysing GARCH–SV convergence, and assessing the volatility effects of COVID-19: evidence from the USA and Japan

---

Lilyves Tchenang & Fátima Lastra Incera

Academic Year: 2025–2026

Abstract

This work investigates several key approaches to estimate Value-at-Risk (VaR). We use parametric, non-parametric, and semi-parametric methods based on Extreme Value Theory (EVT), conducted under both the Block Maxima framework, using the Generalized Extreme Value (GEV) distribution, and the Peaks Over Threshold (POT) framework, using the Generalized Pareto Distribution (GPD). In addition, time-series-based methods are considered, including a Filtered Historical Simulation approach based on volatility smoothing and the GARCH(1,1) model. Empirical results indicate that non-parametric and EVT-based methods generate the most conservative VaR estimates due to their ability to capture heavy-tailed return behavior. Volatility-filtered approaches yield lower and more stable VaR levels, as filtering mitigates short-term volatility clustering. Among time-series models, filtered historical simulation produces the lowest VaR, while GARCH estimates are slightly higher, reflecting their greater sensitivity to recent volatility shocks. The second part of the study examines the theoretical convergence of GARCH(1,1) model to a continuous-time stochastic volatility diffusion. Using simulated high-frequency data, we observe that GARCH parameter estimates exhibit strong persistence and approach the diffusion-limit behavior as sampling frequency increases. However, when the model is applied to real high-frequency financial data, this convergence is weaker, suggesting that market microstructure effects and empirical noise significantly alter the idealized theoretical dynamics. Finally, we analyze the impact of the COVID-19 pandemic on financial market volatility in the USA and Japan during the period 2015–2024. Employing Newey–West heteroskedasticity and autocorrelation consistent (HAC) tests, we find a significant increase in average volatility in the USA market after the onset of the pandemic, while no statistically significant change is detected for Japan. Moreover, although the two markets display different volatility levels in the pre-COVID period, this difference disappears in the post-COVID sample, indicating a convergence in mean volatility across markets following the pandemic shock.

Keywords: Value-at-Risk, Extreme Value Theory, GARCH, Stochastic Volatility, High-Frequency Data, COVID-19, Newey–West Estimator, Bartlett Kernel.

Acronyms

ACF	Autocorrelation Function
ARCH	Autoregressive Conditional Heteroskedasticity
COVID-19	Coronavirus Disease 2019
EVT	Extreme Value Theory
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GEV	Generalized Extreme Value
GPD	Generalized Pareto Distribution
HAR	Heteroskedasticity and Autocorrelation Robust
HAC	Heteroskedasticity and Autocorrelation Consistent
HFD	High-Frequency Data
IID	Independent and Identically Distributed
IMA	Internal Model-based Approach
LRV	Long-Run Variance
MDA	Maximum Domain of Attraction
MLE	Maximum Likelihood Estimator
MRL	Mean Residual Life
P&L	Profit and Loss
POT	Peaks Over Threshold
SMM	Standardized Measurement Method
SV	Stochastic Volatility
VaR	Value-at-Risk
CLT	Central Limit Theorem
CDF	Cumulative Distribution Function
PDF	Probability Density Function
QMLE	Quasi Maximum Likelihood Estimator

Index

1 Introduction

The concept of risk management dates back to the 1950s, following Crockford (1982). Financial institutions can be exposed to five types of risk: market risk, credit risk, counterparty credit risk, liquidity risk and operational risk. In 1996, **market risk** was incroporated into the regulatory framework, as several major financial events. Market risk refers to the risk of potential losses resulting from changes in financial market prices, such as interest rates, stock prices, exchange rates, commodity prices or options prices. Under the Basel I and Basel II Accords, banks must calculate *capital charges* using either the Standardized Measurement Method (SMM) or the Internal Model-based Approach (IMA). It is within the IMA framework that the *Value-at-Risk* (VaR) was introduced.

The VaR can be defined as the potential loss which the portfolio w can suffer for a given confidence level α and a fixed holding period h. In other words, Hull (1989) interprets the VaR as an answer to "How bad can things get?", that is "I am X percent certain there will be a loss of more than Y dollars in the next h days". Roncalli (2020) explains that the VaR can't be computed exactly because the true loss distribution of a portfolio, is unknown and depends on many uncertain and correlated market factors. In this project, we present four approaches to estimate the VaR: (1) the **parametric approach**, (2) the **nonparametric approach** (historical VaR), (3) the **semi-parametric approach** based on **Extreme Value Theory (EVT)** and (4) the **time-series approach**.

In addition, we prove, using simulated and high-frequency data (HFD) that the discrete-time GARCH(1,1) model converges to a continuous-time stochastic volatility diffusion, as showed by Nelson (1990). We also propose a practical procedure to compute VaR from high-frequency returns. Furthermore, we examine the impact of COVID-19 on US and Japanese market volatility (2015–2024), testing for differences before and after the pandemic using the Newey–West estimator with Bartlett's kernel.

2 Testing the effect of pandemic on the volatility in USA and Japan

The outbreak of COVID-19 generated an unprecedented shock to global financial markets, leading to sharp increases in uncertainty and substantial fluctuations in asset prices. In this project, we analyse the impact of the COVID-19 pandemic on stock-market volatility in the USA and Japan. We focus on the S&P 500 and the Nikkei 225 indices over the period 2015–2024, allowing us to compare pre-pandemic market behaviour (2015–2019) with the post-pandemic period (2020–2024) and whether the magnitude of volatility differed between the two markets.

Our time series (X_t^{USA}) and (X_t^{Japan}) are weakly stationary, as squared returns are known to exhibit serial dependence (volatility clustering), violating the i.i.d assumption required for classical t-tests. Therefore, we employ the Newey–West heteroskedasticity- and autocorrelation-consistent (HAC) test, given by (see Appendix A):

$$t = \frac{\bar{Y} - \bar{X}}{\sqrt{\widehat{\mathrm{Var}}(\bar{Y} - \bar{X})}} \xrightarrow{d} \mathcal{N}(0,1), \tag{1}$$

where the variance is estimated using the Bartlett's Kernel:

$$\hat{v} = \sum_{|j| \leq q} w\left(\frac{j}{q}\right) \hat{\gamma}_n(j), \quad w(x) = (1 - |x|)\, \mathbf{1}(|x| < 1). \tag{2}$$

3 Estimating the Value-at-risk

In this section, we present the main methodological approaches for estimating VaR, each relying on different assumptions about the underlying return distribution and the behavior of volatility. The Value-at-Risk cannot be computed directly since the true distribution of portfolio losses is unknown and depends on multiple correlated market factors.

Definition 3.1 (Value at Risk (VaR))**.** Let L be a random variable representing the loss of a portfolio w over a given time horizon T. The **Value at Risk** for a confidence level $\alpha \in (0,1)$ and fixed holding period h, $VaR_\alpha(w;h)$, is the α-quantile of the loss distribution F_L[1]:

$$P(L(w) \leq \text{VaR}_\alpha(w;h)) = \alpha \iff \text{VaR}_\alpha(w;h) = F_L^{-1}(\alpha). \tag{3}$$

If the loss distribution F_L is not continuous, the VaR can be formally defined as:

$$\text{VaR}_\alpha(w;h) = \inf\{x \in \mathbb{R} : \Pr(L(w) \leq x) \geq \alpha\}. \tag{4}$$

According to the Basel Accord I/II, the holding period is set to 10 trading days and the confidence level to 99%. This means that, on average, one *exception* is expected every 100 trading days[2]. Then, the computation of the VaR requires estimating the probability distribution of the portfolio loss, which depends on the underlying market risk factors.

3.1 Parametric estimation or Gaussian VaR

In the parametric estimation approach, also known as the analytical approach, Roncalli (2020) distinguishes two main categories: the **derivation of closed-form formulas**, such as the Gaussian VaR, and the **linear factor models**, which include the covariance model and the factor model. Here, we focus on the Gaussian VaR, which assumes that portfolio losses follow a normal distribution with constant variance.

Let $L_1(w), L_2(w), \ldots, L_n(w)$ be independent and identically distributed (*iid*) realizations (i.e. stationary) of L, such that: $L_i(w) \sim \mathcal{N}(\mu(L), \sigma^2(L))$, for all $i = 1, \ldots, n$. We assume that future losses, L_{n+1}, are independent of past losses and follow the same distribution. Then:

$$P(L_{n+1}(w) \leq \text{VaR}_\alpha(w;h)) = \alpha \iff \Phi\left(\frac{\text{VaR}_\alpha(w;h) - \mu(L)}{\sigma(L)}\right) = \alpha. \tag{5}$$

Hence, the theoretical Gaussian VaR is given by:

$$\frac{\text{VaR}_\alpha(w;h) - \mu(L)}{\sigma(L)} = \Phi^{-1}(\alpha) \iff \text{VaR}_\alpha(w;h) = \mu(L) + \Phi^{-1}(\alpha)\,\sigma(L). \tag{6}$$

In practice, the true mean $\mu(L)$ and volatility $\sigma(L)$ are unknown and must be estimated from the observed sample $L_1, \ldots, L_n$ using maximum likelihood estimators (MLEs). Substituting these estimates into (6), the parametric estimator of VaR becomes[3]:

$$\widehat{\text{VaR}}_\alpha(w;h) = \hat{\mu}(L) + \Phi^{-1}(\alpha)\,\hat{\sigma}(L). \tag{7}$$

[1] Equivalenty, if $\Pi(w)$ denotes the profit and loss (P&L) of the portfolio w, then the VaR can equivalently be expressed in terms of the distribution of $\Pi(w)$: $\Pr\left(\Pi(w) \geq -\text{VaR}_\alpha(w;h)\right) = \alpha$ and $\text{VaR}_\alpha(w;h) = -F_\Pi^{-1}(1-\alpha)$ where F_Π denotes the cumulative distribution function (CDF) of the portfolio P&L.

[2] An exception occurs when the actual loss on a given day exceeds the VaR estimate for that day.

[3] Roncalli (2020) notes that it is very difficult to estimate the mean in practice, so we set $\hat{\mu}(L) = 0$.

3.2 Nonparametric estimation or historical VaR

The historical VaR is a non-parametric estimate of the VaR. For that, we consider the empirical distribution of the risk factors observed in the past. Let $(F_{1,s}, \ldots, F_{m,s})$ be the vector of market risk factors observed at time $s < t$. The P&L of a portfolio with weights w under a historical scenario s can be computed via a pricing function $g(\cdot)$ as: $\Pi_s(w) = g(F_{1,s}, \ldots, F_{m,s}; w) - P_t(w)$, where $P_t(w)$ represents the current portfolio value. Assuming n_S historical observations, the empirical distribution of the P&L can be expressed as a discrete distribution:

$$\widehat{F}_{\Pi}(x) = \frac{1}{n_S} \sum_{s=1}^{n_S} \mathbf{1}(\Pi_s(w) \leq x)\,, \tag{8}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Then, the VaR is estimated as:

$$\widehat{\mathrm{VaR}}_{\alpha}(w; h) = -\widehat{F}_{\Pi}^{-1}(1-\alpha), \tag{9}$$

In practice, the quantile $\widehat{F}_{\Pi}^{-1}(1-\alpha)$ can be obtained using: the **order statistic approach** or the **kernel approach**. Here, we focus on the former. Let $X_1, \ldots, X_n$ be *iid* as X with a continuous distribution F. Suppose that for a given scalar $\alpha \in (0,1)$, there exists a sequence $\{a_n\}$ such that $\sqrt{n}(a_n - n\alpha) \to 0$. According to Lehmann (1998), we have:

$$\sqrt{n}\big(X_{(a_n:n)} - F^{-1}(\alpha)\big) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{\{f(F^{-1}(\alpha))\}^2}\right). \tag{10}$$

with $f(\cdot)$ the density of F. This result implies that the quantile $F^{-1}(\alpha)$ can be estimated by the mean of the $n\alpha$-th order statistic. Applying this result to the P&L sample $\{\Pi_1(w), \ldots, \Pi_{n_S}(w)\}$, we compute the order statistics $\min_s \Pi_s(w) = \Pi_{(1:n_S)} \leq \Pi_{(2:n_S)} \leq \cdots \leq \Pi_{(n_S:n_S)} = \max_s \Pi_s(w)$. Then, by definition we have that $\widehat{F}_{\Pi}(\Pi_{(m)}) = \frac{m}{n_S}$, and if we choose $m = \lfloor n_S\alpha \rfloor$, we obtain $\widehat{F}_{\Pi}(\Pi_{(m)}) \approx \alpha$. Hence, the m-th order statistic $\Pi_{(m)}$ provides a nonparametric estimator of the quantile $F_{\Pi}^{-1}(\alpha)$. Therefore, the VaR at confidence level $1-\alpha$ is given by:

$$\widehat{\mathrm{VaR}}_{\alpha}(w; h) = \begin{cases} -\,\Pi_{(n_S(1-\alpha):n_S)}(w), & \text{if } n_S(1-\alpha) \in \mathbb{N}, \\ -\Big[\Pi_{(q:n_S)}(w) + (n_S(1-\alpha) - q)\big(\Pi_{(q+1:n_S)}(w) - \Pi_{(q:n_S)}(w)\big)\Big], & \text{if } n_S(1-\alpha) \notin \mathbb{N}, \end{cases}$$

where $q = \lfloor n_S(1-\alpha) \rfloor$.

3.3 Extreme value theory

The Extreme Value Theory focuses on modelling the behaviour of the *tails* of a distribution, that is, the occurrence of rare and extreme losses. It studies the limit distribution of extreme order statistics such as the sample minimum $X_{1:n}$ and the sample maximum $X_{n:n}$ as $n \to \infty$. There are two main parametric approaches: the **Block Maxima** approach, which leads to the **Generalized Extreme Value** distribution and the **Peaks Over Threshold (POT)** approach, which leads to the **Generalized Pareto Distribution**.

3.3.1 Block maxima approach

The Block Maxima approach consists on dividing a time series of observations into non-overlapping blocks of equal size, and then to extract the maximum (or minimum) value within

each block to model the tail behaviour of the distribution of extreme events[4].

First, we define the Maximum Domain of Attraction (MDA) of a distribution, which indicates the type of extreme value distribution to which the normalized maxima of a distribution converge. Then, we present the **Fisher–Tippett Theorem**, which provides a universal characterization of the limiting behaviour of extreme events. It states that, under general conditions, the distribution of maxima (or minima) of independent random variables converges to one of three types of extreme value distributions: Gumbel, Fréchet, or Weibull. In this sense, the theorem plays a role analogous to the Central Limit Theorem (CLT).

Definition 3.2 (Maximum Domain of Attraction (MDA))**.** Suppose that $X_1, \ldots, X_n$ is a sequence of *i.i.d.* random variables with cumulative distribution function F. Let $M_n = \max\{X_1, \ldots, X_n\}$. Then, F is said to be in the **maximum domain of attraction** of G, $MDA(G)$. If there exist sequences of constants $a_n > 0$ and b_n such that, as $n \to \infty$,

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \longrightarrow G(x) \tag{11}$$

for some non-degenerate distribution function G.

Theorem 3.1 (Fisher-Tippet Theorem)**.** *If $F \in MDA(G)$ for some non-degenerate distribution function G, then G must be of the same type as one of the extreme value distributions:*

- ***Gumbel (Type I):*** $G(x) = \exp\{-e^{-x}\}, \quad -\infty < x < \infty.$
- ***Fréchet (Type II):*** $G(x) = \begin{cases} \exp(-x^{-\alpha}), & x > 0, \\ 0, & x \leq 0, \end{cases} \qquad \alpha > 0.$
- ***Negative Weibull (Type III):*** $G(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x < 0, \\ 1, & x \geq 0, \end{cases} \qquad \alpha > 0.$

The three types can be combined into the **GEV** distribution, whose CDF is the following[5]:

$$G_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left\{-(1+\xi\left(\frac{x-\mu}{\sigma}\right))^{-1/\xi}\right\}, & \xi \neq 0; \text{where } 1+\xi x > 0, \\ \exp\left\{-e^{-\left(\frac{x-\mu}{\sigma}\right)}\right\}, & \xi = 0, \end{cases} \tag{12}$$

with μ the **location parameter**, σ the **scale parameter**, and ξ the **shape parameter**.

Now, let us apply this method to the estimation of VaR. Let us consider a portfolio w with a mark-to-market value $P_t(w)$ at time t. The P&L of the portfolio between two consecutive trading days, t and $t+1$, can be expressed as $\Pi(w) = P_{t+1}(w) - P_t(w) = P_t(w) \cdot R(w)$, where $R(w)$ denotes the daily return of the portfolio. If we denote by $\widehat{F}_R$ the estimated CDF of returns $R(w)$, the Value-at-Risk at confidence level α can be expressed as:

$$\widehat{\mathrm{VaR}}_\alpha(w) = -P_t(w) \cdot \widehat{F}_R^{-1}(1-\alpha), \tag{13}$$

where the negative sign accounts for the fact that losses correspond to negative returns.

[4]There is a trade-off in choosing block size: larger blocks yield more extreme maxima but fewer observations, while smaller blocks give more data but weaker extremes. In financial applications, using monthly blocks (20 trading days) is a common consensus.

[5]Then, we have Gumbel $\sim$ GEV$(0,1,0)$, Fréchet $\sim$ GEV$(1,\alpha^{-1},\alpha^{-1})$, and Negative Weibull $\sim$ GEV$(-1,\alpha^{-1},-\alpha^{-1})$.

Then, we estimate the GEV distribution $\widehat{G}$ of the maxima of the negative returns $-R(w)$ over blocks of n trading days. Let α_{GEV} denote the confidence level associated with the GEV distribution fitted to the block maxima. To ensure consistency between the daily and block-based confidence levels, we impose that both represent the same return period (i.e., the same expected frequency of exceedances) and we deduce that $\alpha_{\text{GEV}} = 1 - n(1-\alpha)$, which allows us to compute the block-level confidence level α_{GEV} corresponding to a given daily confidence level α. Hence, the VaR using the estimated GEV distribution is given by:

$$\widehat{\text{VaR}}_\alpha(w;h) = P_t(w) \cdot \widehat{G}^{-1}(\alpha_{\text{GEV}}). \tag{14}$$

3.3.2 Peaks Over Threshold

Roncalli (2020) explains that the previous method can lead to a loss of information, as some blocks may contain multiple extreme events, while others may include none at all. To address this limitation, the **Peaks Over Threshold** approach models all observations that exceed a predefined high threshold (see Appendix C). Then, according to the **Pickands–Balkema–de Haan Theorem**, for a sufficiently high threshold, the distribution of exceedances can be well approximated by the Generalized Pareto Distribution.

Definition 3.3 (Excess distribution function)**.** Let X be a random variable with distribution function F and right endpoint $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. For fixed $0 < u < x_F$, we define the conditional distribution of exceedances over u for $0 \le x < x_F - u$ as:

$$F_u(x) = \Pr\{X - u \le x \mid X > u\} = \frac{P(u < X \le u + x)}{P(X > u)} = \frac{F(u+x) - F(u)}{1 - F(u)}. \tag{15}$$

Theorem 3.2 (Pickands–Balkema–de Haan Theorem)**.** *Let F_u be an excess distribution. If F_u belongs to the MDA of an extreme value distribution with shape parameter ξ, then there exists a positive scaling function $\sigma(u)$ such that:*

$$\lim_{u \to x_F} \sup_{0 \le x < x_F - u} \left| F_u(x) - H_{\xi,\sigma(u)}(x) \right| = 0, \tag{16}$$

where $H_{\xi,\sigma}(x)$ denotes the GPD, whose CDF is the following:

$$H_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \xi\left(\frac{x}{\sigma}\right))^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-x/\sigma}, & \xi = 0, \end{cases} \tag{17}$$

*where $\sigma > 0$ is the **scale parameter** and ξ is the **shape parameter**.*

Now, let's see how to apply this result to estimating the VaR. We havce that $F_u(x) \sim H(x) \sim \text{GPD}(\sigma, \xi)$. This way, for $x \ge u$, $F(u+x) = F(u) + \big(1 - F(u)\big)F_u(x)$ and redefining $x = x + u$, we obtain $F(x) = F(u) + \big(1 - F(u)\big)\, F_u(x-u) \;\approx\; F(u) + \big(1 - F(u)\big)\, H(x-u)$. Given a sample of size n, let n_u be the number of observations exceeding u, the nonparametric estimate of $F(u)$ is $\widehat{F}(u) = 1 - n_u/n$. Therefore, we obtain:

$$\widehat{F}(x) = \widehat{F}(u) + \big(1 - \widehat{F}(u)\big)\, \widehat{H}(x-u) = 1 \;-\; \frac{n_u}{n}\left(1 + \widehat{\xi}\,\frac{x-u}{\widehat{\sigma}}\right)^{-1/\widehat{\xi}}. \tag{18}$$

Hence, the POT estimator of the VaR is:

$$\widehat{\mathrm{VaR}}_\alpha(w;h) \;=\; \widehat{F}^{-1}(\alpha) \;=\; u \;+\; \frac{\widehat{\sigma}}{\widehat{\xi}}\left[\left(\frac{n}{n_u}(1-\alpha)\right)^{-\widehat{\xi}} - 1\right]. \tag{19}$$

3.4 Time series models

In this section, we present the estimation of VaR using time series models. There are two main approaches to estimate the VaR: (1) **Filtered time series** and the **GARCH model**.

3.4.1 Filtered time series

The Filtered Time Series approach estimates the VaR of a process X_t by first removing the non-stationarity of returns through volatility filtering. This produces standardized residuals that are approximately independent, whose empirical quantiles are then used to compute the VaR, subsequently rescaled by the forecasted volatility for the next period.

Definition 3.4 (Filtered Time Series)**.** Let $\{X_t\}_{t=1}^n$ denote a time series with non-stationary behavior in its mean and volatility, then we define the *filtered time series* $\{Z_t\}$ as:

$$Z_t = \frac{X_t - \widetilde{E}(X_t)}{\widetilde{SD}(X_t)}, \tag{20}$$

where $\widetilde{E}(X_t)$ denotes the estimated **locally smoothed mean** of X_t and $\widetilde{SD}(X_t)$ denotes the estimated **locally smoothed standard deviation** of X_t.

Suppose we have a time series of length T, where $X_t(w)$ denotes the **return** of the portfolio at time t. The VaR for day $T+1$ can then be estimated as:

$$\widehat{\mathrm{VaR}}_\alpha(w;T+1) = \mathrm{VaR}_Z(\alpha) \times \widetilde{SD}_f(X_t(w)) + \widetilde{E}_f(X_t(w)). \tag{21}$$

3.4.2 GARCH

In 1982, Engle (1982) introduced the *Autoregressive Conditional Heteroskedasticity* (ARCH) model to account for time-varying volatility in financial series. However, ARCH models require a large number of parameters since the conditional variance depends solely on past squared shocks. To overcome this limitation, Bollerslev (1986) proposed the *Generalized ARCH* (GARCH) model, which allows the conditional variance to depend not only on past squared innovations but also on its own past values. Here, we focus on the standard GARCH(1,1) specification, which remains one of the most widely used models for volatility estimation.

Definition 3.5 (GARCH(1,1))**.** Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a stochastic process. The process is said to follow a *Generalized Autoregressive Conditional Heteroskedasticity model of order (1,1)*, denoted as GARCH(1,1), if it satisfies the following conditions:

1. (ε_t) is (strictly) stationary;
2. $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$, where $\mathcal{F}_{t-1} = \sigma(\varepsilon_s,\, s \le t-1)$;
3. $\varepsilon_t = \sigma_t Z_t$, where $(Z_t)_t$ is an i.i.d. sequence with mean 0 and variance 1;
4. $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, with $\omega > 0$, $\alpha \ge 0$, and $\beta \ge 0$.

The VaR can be computed directly from a GARCH model. Let the return process (r_t) follow a GARCH(1,1) model defined as $r_t = \mu_t + \varepsilon_t$. At time t, the one-step-ahead forecast for

the conditional mean and variance is given by μ_{t+1} and σ^2_{t+1}, respectively. The VaR can be interpreted as the conditional quantile of the future shock ε_{t+1} given the available information $\mathcal{F}_t$. Formally, we seek the value v such that:

$$\mathbb{P}(\varepsilon_{t+1} \leq v \mid \mathcal{F}_t) = \alpha. \tag{22}$$

Since $\varepsilon_{t+1} = \sigma_{t+1} Z_{t+1}$, this is equivalent to $\mathbb{P}\left(Z_{t+1}\sigma_{t+1} \leq v\right) = \mathbb{P}(Z_{t+1} \leq z_\alpha) = \alpha$. It then follows that the conditional quantile of ε_{t+1} is $v = \sigma_{t+1} z_\alpha$. Therefore, the conditional VaR at confidence level α for day $t+1$ is computed as:

$$\widehat{\mathrm{VaR}}_\alpha(w; t+1) = \mu_{t+1} + \sigma_{t+1} z_\alpha, \tag{23}$$

where z_α denotes the α-quantile of the standard normal distribution. However, we need to estimate σ_{t+1}, so we need to estimate $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$. Assuming that we observe a sample $\varepsilon_1, \ldots, \varepsilon_n$ and an initial value for the variance σ_1^2, the likelihood and log-likelihood functions are given by:

$$\begin{cases} L(\theta) = f(\varepsilon_1, \ldots, \varepsilon_n; \theta) = f(\varepsilon_1) \prod_{t=2}^n f(\varepsilon_t \mid \varepsilon_{t-1}, \sigma_{t-1}) = f(\varepsilon_1) \prod_{t=2}^n \frac{1}{\sigma_t} \phi\left(\frac{\varepsilon_t}{\sigma_t}\right), \\ \ell(\theta) = \log f(\varepsilon_1) + \sum_{t=2}^n \ell_t(\theta), \quad \text{with} \quad \ell_t(\theta) = \log f(\varepsilon_t \mid \varepsilon_{t-1}, \sigma_{t-1}), \end{cases} \tag{24}$$

where $\theta = (\omega, \alpha, \beta)$ denotes the parameter vector to be estimated. In practice, the first conditional variance σ_1^2 is not observed, so we replace it by an empirical estimate. This value is the starting point for the recursion of σ_t^2 in the likelihood computation and the final estimates $\widehat{\omega}$, $\widehat{\alpha}$, and $\widehat{\beta}$ are obtained using the quasi-maximum likelihood estimator (QMLE).

4 GARCH and Stochastic Volatility Models

The most popular approach to model the volatility smile is to consider that the volatility is not constant, but stochastic. In the stochastic volatility (SV) framework, volatility is treated as a latent and time-varying process that cannot be directly observed. The asset price and its volatility are jointly driven by a system of stochastic differential equations:

$$\begin{cases} dX_t = \mu X_t \, dt + f(\sigma_t) \, X_t \, dW_t^{(1)}, \\ d\sigma_t = a(\sigma_t) \, dt + b(\sigma_t) \, dW_t^{(2)}, \end{cases} \tag{25}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are standard Brownian motions. The process $(X_t)_{t \geq 0}$ is theoretically defined in continuous time, but, in practice, we only observe a finite number of discrete samples $0 = t_0 < t_1 < \cdots < t_n = 1$. If we observe n values of the price over one day, the sampling step is $\Delta t = 1/n$. Hence, the available data consist of discrete observations $(X_{i/n})_{i=0,\ldots,n}$ and if n is large, then we are in a **high-frequency setting**.

Wang (2002) explain that continuous-time models have been widely used in theoretical finance, especially for option pricing, but as in practice we only observe data at discrete time points, discrete-time models like GARCH are commonly used when working with real financial data. Recently, several studies have tried to connect both approaches by showing how GARCH models can converge to stochastic volatility models in continuous time.

Nelson (1990) is the first one to proof that, under high-frequency sampling and appropriate parameter scaling, the discrete-time GARCH(1,1) model converges to a continuous-time SV diffusion. In other words, when the observation interval becomes very small, the GARCH process behaves like a SV model in continuous time, with the conditional variance evolving

randomly as in an SV diffusion. This is expressed in the following Garch-SV model (see Appendix F):

$$\begin{cases} dX_t = \sigma_t \, dW_t^{(1)}, \\ d\sigma_t^2 = (\omega_0 - \beta_0 \sigma_t^2) \, dt + \sqrt{2\alpha_0} \, \sigma_t^2 \, dW_t^{(2)}, \end{cases} \tag{26}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are independent, $\omega_0 > 0$, $\beta_0 > 0$, and $\alpha_0 \in (0, 1)$. In SV models, volatility is unobserved and the likelihood is hard to compute. A practical approach is to estimate the GARCH(1,1) parameters (ω, α, β) via MLE, and then translate these estimates into the corresponding continuous-time parameters $(\omega_0, \alpha_0, \beta_0)$.

In this project, we investigate this theoretical link on both simulated data and real high-frequency returns to assess whether this theoretical relationship remains valid in practice and to evaluate the suitability of GARCH models for high-frequency VaR estimation. All simulation details and methodological considerations are presented in Appendix G.

4.1 Computation of VaR with high frequency time series data

Using the GARCH(1,1) specification introduced in 2.4.2, we adapt the model to the high-frequency setting. Tick-by-tick returns are typically very small, exhibit strong volatility clustering, and have negligible conditional mean. For this reason, once the GARCH parameters have been estimated from high-frequency data, the one-step-ahead volatility forecast obtained from the model can be directly used to compute intraday risk measures.

Let (X_t) denote the high-frequency log-returns. Since $\mu_t \approx 0$ at very short horizons, the GARCH(1,1) forecast σ_{t+1} derived earlier simplifies the VaR computation. In particular, because the next-period innovation satisfies: $\varepsilon_{t+1} = \sigma_{t+1} Z_{t+1}$, $Z_{t+1} \sim \mathcal{N}(0, 1)$, the conditional α-quantile of X_{t+1} reduces to a rescaling of the normal quantile:

$$\mathrm{VaR}_X(\alpha)[t+1] = \sigma_{t+1} \, z_\alpha. \tag{27}$$

This formulation yields a fully dynamic intraday VaR series: as each new high-frequency return arrives, the GARCH recursion updates σ_{t+1} at each tick, and the VaR adjusts immediately to reflect rapid changes in volatility, driven by market microstructure effects

5 Results and Discussion

In this section, we present and discuss the empirical findings of the study.

In Table 1 compares the results obtained from the different VaR estimation methods, along with their corresponding confidence intervals (see Appendix D). First, non-parametric and EVT-based approaches[6] produce the highest VaR levels, with values ranging from approximately. This reflects their ability to better capture heavy-tailed behaviour in financial returns, leading to more conservative risk estimates compared to the parametric Gaussian-based approach. Second, the filtered versions of all models yield substantially lower VaR estimates. This reduction is expected, as volatility filtering removes short-term volatility clustering and produces residuals with thinner tails, resulting in less extreme quantiles. Consequently, filtered VaR tends to be more stable over time. Third, the filtered VaR delivers the lowest VaR estimate. In contrast, the GARCH-based VaR is slightly higher, reflecting the greater responsiveness of the GARCH conditional variance to recent volatility shocks.

[6]EVT models full diagnostic analysis is presented in Appendix E.

Overall, the methods can be ranked by VaR magnitude as follows: Block Maxima, Non-parametric, POT, Parametric, Filtered, and GARCH, which produces the lowest estimates.

Table 1: Comparison of Value-at-Risk estimation methods

Method	**VaR**	**CI95 (VaR)**	**Filtered VaR**	**CI95 (Filtered)**
Parametric	0.026569	[0.024991, 0.028361]	0.021094	[0.019809, 0.022555]
Non-parametric	0.031332	[0.022651, 0.032469]	0.026282	[0.018414, 0.026672]
Block maxima	0.032378	[0.021923, 0.053902]	0.022794	[0.017710, 0.028204]
POT	0.029239	[0.024743, 0.033391]	0.023843	[0.017633, 0.022599]
Filtered	0.024669	[0.018414, 0.026672]	-	-
Garch	0.025092	[0.023601, 0.026784]	-	-

Table 2 reports the GARCH(1,1) parameter estimates on simulated data for each subsampling rate with baseline parameters $\omega_0 = 0.02$, $\alpha_0 = 0.3$, $\beta_0 = 1.5$. In Table 2 we can see that the sum $\hat{\alpha} + \hat{\beta}$ is close to 1 for most subsampling rates, particularly for $m = 10$ and $m = 50$, which is consistent with the theoretical diffusion-limit behaviour of high-frequency GARCH: as sampling frequency increases the discrete-time estimates show strong persistence and tend toward the continuous-time limit. For very high frequency (rate 1) we observe $\hat{\alpha} \approx 0.28$, $\hat{\beta} \approx 0.67$ while for intermediate rates we obtain smaller $\hat{\alpha}$ and larger $\hat{\beta}$ with persistence near 1.

Table 2: GARCH(1,1) estimates on subsampled simulated data.

Subsampling rate m	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
1	0.00000	0.12240	0.87760	1.00000
5	0.00000	0.17491	0.81978	0.99469
10	0.00000	0.22222	0.76633	0.98855
50	0.00000	0.28350	0.71556	0.99905
100	0.00000	0.42333	0.57538	0.99871

Table 3 reports the parameter estimates obtained for the real data for different subsampling rates. Table 3 shows the persistence measure $\hat{\alpha} + \hat{\beta}$ is significantly lower than what we obtained on simulated data, suggesting much weaker clustering of volatility than predicted by the theoretical diffusion model. This highlights that while GARCH(1,1) converges nicely toward the diffusion limit when the data truly follows an SV model, real *market microstructure* strongly perturbs this theoretical behaviour.

Table 3: Estimated GARCH(1,1) parameters on sumsampled real data.

Subsampling rate m	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
1	0.00000	0.02642	0.47551	0.50192
5	0.00000	0.23312	0.58679	0.81991
10	0.00000	0.09794	0.51602	0.61397
50	0.00000	0.16092	0.50466	0.66557
100	0.00000	0.02631	0.47350	0.49980

After fitting the GARCH(1,1) model to the full high-frequency return series, we compute the one-step-ahead VaR, using the estimated parameters $\widehat{\omega} = 0.00000$, $\widehat{\alpha} = 0.02642$, and $\widehat{\beta} = 0.47551$. The first five VaR values obtained are:

$$\text{VaR}_{1:5} = (-0.00060251,\ -0.00062122,\ -0.00062992,\ -0.00063401,\ -0.00073043).$$

Table 4 summarizes the HAC t-test results for the equality of mean volatility across periods and countries. First, in Table 4 we see that the null hypothesis for the USA is strongly rejected, indicating a statistically significant different volatility in the post-COVID period. In contrast, the null cannot be rejected for Japan, suggesting that the average level of volatility in the Japanese market did not experience a significant shift between the pre- and post-COVID periods. Then, in Table 4, the USA and Japan exhibit significantly different levels of mean volatility in the pre-COVID period, but this difference disappears in the post-COVID period, where the null hypothesis is not rejected. This result implies a convergence in mean volatility between the two markets after the onset of the pandemic.

Table 4: HAC t-tests for equality of mean volatility.

Hypothesis	**p-value**	**Result**
$H_0 : \mu_{\text{USA, pre}} = \mu_{\text{USA, post}}$	0.0000	Reject H_0
$H_0 : \mu_{\text{JPN, pre}} = \mu_{\text{JPN, post}}$	0.0836	Do not reject H_0
$H_0 : \mu_{\text{USA, pre}} = \mu_{\text{JPN, pre}}$	0.0000	Reject H_0
$H_0 : \mu_{\text{USA, post}} = \mu_{\text{JPN, post}}$	0.7900	Do not reject H_0

6 Conclusions

This project has studied four approaches to estimate the VaR: (1) the parametric approach, (2) the nonparametric approach (historical VaR), (3) the semi-parametric approach based on Extreme Value Theory (EVT), and (4) the time-series approach. Non-parametric and EVT-based methods yield the most conservative VaR estimates by capturing heavy-tailed return behavior, whereas filtered approaches produce lower and more stable VaR levels by mitigating short-term volatility clustering. Among time-series models, filtered VaR delivers the lowest VaR, while GARCH-based VaR is slightly higher due to its greater responsiveness to recent volatility shocks.

The theoretical link between discrete-time GARCH models and continuous-time stochastic volatility diffusions was also investigated. Using simulated high-frequency data, GARCH(1,1) parameter estimates exhibit strong persistence and approach the diffusion-limit behaviour predicted by theory as sampling frequency increases. However, when applied to real high-frequency data, this convergence weakens substantially, suggesting that market microstructure effects and empirical noise significantly distort the idealized theoretical dynamics.

Finally, the analysis of volatility during the COVID-19 period reveals clear structural differences across markets. While U.S. equity volatility increased significantly in the post-pandemic period, no comparable statistically significant shift is detected for Japan. Moreover, although the two markets displayed distinct volatility levels prior to COVID-19, these differences largely vanish after the pandemic, indicating a convergence in mean volatility across markets during periods of global stress.

References

Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econom. Rev.*, 5(1):1–50.

Crockford, G. N. (1982). The bibliography and history of risk management: Some preliminary observations. *Geneva Pap. Risk Insur. Issues Pract.*, 7(2):169–179.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007.

Hull, J. C. (1989). *Options, Futures, and Other Derivatives*. Pearson, Upper Saddle River, NJ, 8 edition.

Kolokotrones, T., Stock, J. H., and Walker, C. D. (2023). Is Newey–West optimal among first-order kernels? *J. Econom.*, (105399):105399.

Lehmann, E. L. (1998). Nonparametric estimation. In *Elements of Large-Sample Theory*, pages 451–570. Springer-Verlag, New York.

Naess, A. (2024). The peaks-over-threshold method. In *Applied Extreme Value Statistics*, pages 19–28. Springer Nature Switzerland, Cham.

Nelson, D. B. (1990). ARCH models as diffusion approximations. *J. Econom.*, 45(1-2):7–38.

Roncalli, T. (2020). *Handbook of financial risk management*. Chapman and Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Philadelphia, PA.

Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Stat.*, 30(3):754–783.

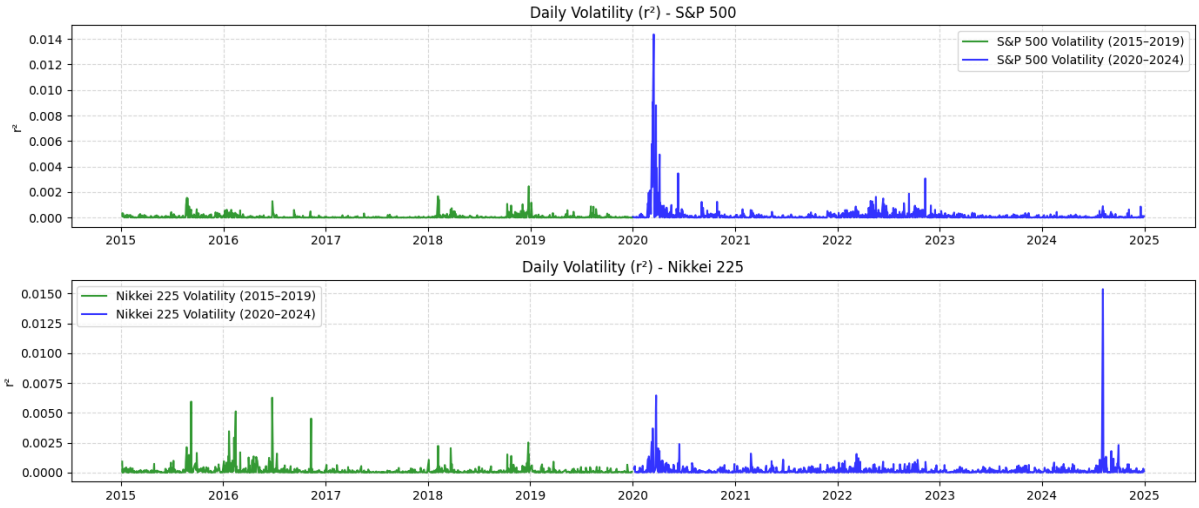
Appendix A: Statistical inference under serial correlated time series

In this appendix, we develop the theoretical tools required to perform valid statistical inference on our financial volatility time series in order to test whether average market volatility changed after the COVID–19 pandemic and whether the volatility of the S&P 500 and the Nikkei 225 differed across the pre- and post-pandemic periods. We focus on the estimation of the **mean** and the **covariance structure** of a **weakly stationary** time series, as these two elements together characterise its fundamental statistical behaviour. The mean represents the long-run average level of the process (in our case, the average daily volatility), while the covariance structure captures the temporal dependence between observations.

Specifically, we study the daily volatility series (squared returns) of the S&P 500 and Nikkei 225, indices over the periods 2015–2019 (pre-COVID) and 2020–2024 (post-COVID). We want to test whether the pre-pandemic volatility is the same as the post-pandemic period and whether the magnitude of volatility differed between the two markets in each period.

Figure 1 presents the daily volatility series (r_t^2) of the S&P 500 and Nikkei 225 indices over the periods 2015–2019 (green) and 2020–2024 (blue). We observe that both markets exhibit a marked increase in volatility during the post-COVID period, with several pronounced spikes. In contrast, the pre-COVID period shows relatively lower and more stable levels of volatility, particularly for the S&P 500.

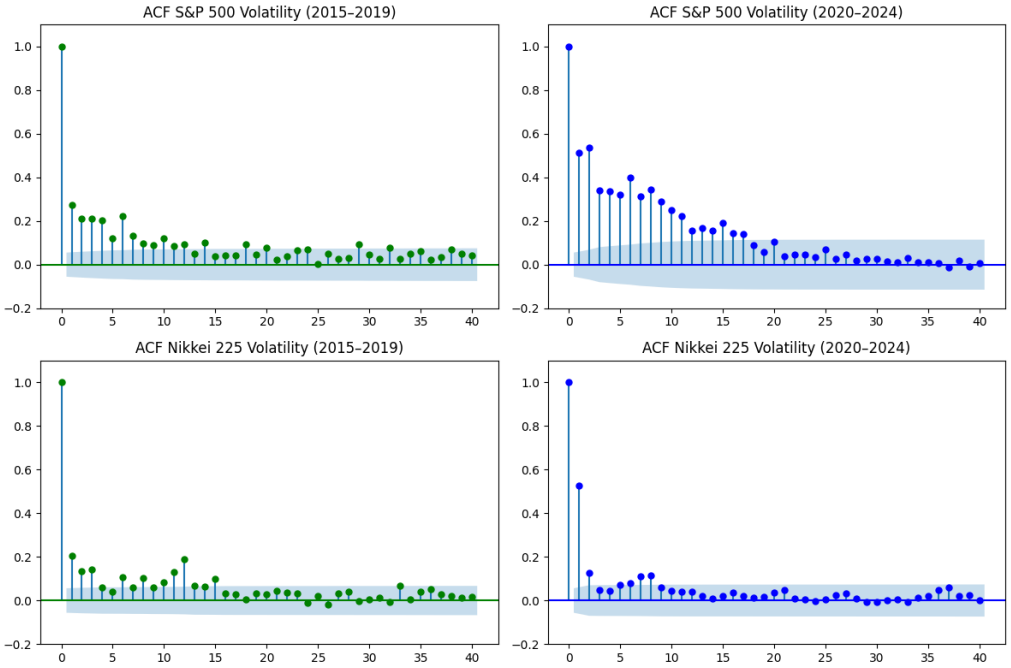
Figure 1: Daily volatility series (r_t^2) of the S&P 500 and Nikkei 225 indices over the periods 2015–2019 (pre-COVID) and 2020–2024 (post-COVID).



First, we analyze our time series and find that both can be considered **weakly stationary**, as their mean, variance, and autocovariance remain constant over time. However, the series exhibit strong serial dependence, meaning that the usual assumption of i.i.d. observations does not hold, as seen in Figure 2, which displays the sample autocorrelation functions (ACFs) of daily volatility for both indices. The ACFs reveal strong and persistent serial correlation, which gradually decays as the lag increases. This pattern confirms the presence of **volatility clustering**, a known fact in financial time series, where periods of high volatility tend to be followed by high volatility, and periods of low volatility by low volatility.

Consequently, classical inference procedures, such as the standard t-test for equality of means, would underestimate the true variability of the sample mean. This appendix develops the theoretical framework required to conduct valid inference under temporal dependence.

Figure 2: Sample autocorrelation functions (ACFs) of daily volatility (r_t^2) for the S&P 500 and Nikkei 225 indices over the periods 2015–2019 (pre-COVID) and 2020–2024 (post-COVID).



Let us consider a time serie $\{X_t\}_{t=1}^n$ of n daily volatility observations $(X_1, X_2, \ldots, X_n)$ for a specific period. If the volatility observations were i.i.d., then the variance of the sample mean would be the usual sampling variance. However, financial volatility data presents **serial dependence** and **conditional heteroskedasticity** (**volatility clustering**). As a result, the usual sampling variance no longer applies, and we must rely on the following result.

Proposition 6.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly stationary time series with mean μ, autocovariance function $\gamma(\cdot)$ and autocorrelation function $\rho(\cdot)$. Then, as $n \to \infty$:*

1. *If $\rho(h) \to 0$ as $h \to \infty$, then $\bar{X}_n \xrightarrow[n\to\infty]{L^2} \mu$.*
2. *If $\sum_h |\rho(h)| < \infty$, then $n\,\mathrm{Var}(\bar{X}_n) \xrightarrow[n\to\infty]{} v$, where $v = \sum_{h\in\mathbb{Z}} \gamma(h)$.*
3. *If $\sum_h |\rho(h)| < \infty$ and some additional "mixing condition" holds, then the CLT applies $\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, v)$, where the long-run variance (LRV) is defined as:*

$$v = \sum_{h=0}^{\infty} \gamma(h) = \gamma(0) + 2\sum_{h=1}^{\infty} \gamma(h) = \mathrm{Var}(X) + 2\sum_{h=1}^{\infty} \mathrm{Cov}(X_0, X_h). \tag{28}$$

If $\gamma(h) > 0$ for $h \geq 1$, then the variance of the sample mean is larger than in the i.i.d. case and viceversa.

From this proposition, we can extract the following conclusions. First, although the volatility series are serially correlated, the sample mean over time remains a consistent estimator of the true average volatility within each period. Second, the variability of the sample mean depends not only on the individual variance $\gamma(0)$, but on all the autocovariances $\gamma(h)$ at different lags. Third, when the data are correlated, the variance in the CLT becomes the LRV. In practice, this LRV is estimated using the Newey–West estimator, which provides a consistent estimate of v even when the data are autocorrelated or conditionally heteroskedastic.

Definition 6.1 (Newey–West estimator)**.** Let (X_t) be a weakly stationary process with autocovariance function $\gamma(h)$. Then, the LRV is estimated by a weighted sum of sample covariances:

$$\hat{v} = \sum_{|j|\leq q} w\left(\frac{j}{q}\right)\hat{\gamma}_n(j), \tag{29}$$

with $w(\cdot)$ the kernel function satisfying $w(0) = 1$, $q = q_n$ is the *bandwidth* such that $q \to \infty$ and $q/n \to 0$ as $n \to \infty$, and $\hat{\gamma}_n(h)$ is the empirical autocovariance function, given by:

$$\hat{\gamma}_n(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n), \quad \forall h \geq 0 \tag{30}$$

Kolokotrones et al. justify that, although no single first-order kernel is theoretically optimal for heteroskedasticity and autocorrelation robust (HAR) testing, none of the alternative first-order positive–semidefinite kernels outperform the **Bartlett kernel**.

Definition 6.2 (Bartlett kernel)**.** The Bartlett kernel or triangular kernel, is a first-order kernel function commonly used in the estimation of the LRV within HAC estimators. It is defined as:

$$w(x) = (1 - |x|)\,\mathbf{1}(|x| < 1). \tag{31}$$

The hypothesis of equal mean volatility between two periods or markets is tested using a HAR t-statistic, based on the Newey–West heteroskedasticity and autocorrelation consistent (HAC) covariance estimator:

$$t = \frac{\bar{Y} - \bar{X}}{\sqrt{\widehat{\mathrm{Var}}(\bar{Y} - \bar{X})}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{32}$$

Therefore, we reject the null hypothesis if $|t| > z_{1-\alpha/2}$, or, equivalently, if the corresponding p-value is smaller than the chosen significance level.

Appendix B: Duality between GEV and GPD distribution

In this appendix, we present one interesting result in EVT that is required in order to properly understand the threshold selection procedure for the POT approach explained in Appendix C: the duality between the GEV and GPD distributions. There are several ways to characterize this, but we focus on the approach based on the extremal limit distribution and its Taylor expansion, as Roncalli (2020).

Assuming that block maxima follow a GEV(μ, σ, ξ) distribution, the cumulative distribution function of the maximum can be approximated as:

$$F^n(x) \approx \exp\left[-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right]. \tag{33}$$

Taking logarithms on both sides of the equation (33) and using the taylor approximation, $\ln F(x) \approx -(1 - F(x))$ for large x, we derive:

$$1 - F(x) \approx \frac{1}{n}\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}. \tag{34}$$

Now, consider exceedances above a high threshold u. The conditional tail distribution satisfies:

$$\Pr(X > u + x \mid X > u) = \frac{1 - F(u + x)}{1 - F(u)}. \tag{35}$$

Substituting (34) into (35), we obtain the GPD:

$$\Pr(X > u + x \mid X > u) = \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-1/\xi}, \tag{36}$$

with adjusted scale parameter $\tilde{\sigma} = \sigma + \xi(u - \mu)$. Therefore, exceedances above a high threshold u follow a GPD$(\tilde{\sigma}, \xi)$ distribution, linking the POT framework and the GEV distribution.

Appendix C: Threshold selection for the POT approach

In this appendix, we discuss the procedures used to select an appropriate threshold for the POT approach. The estimation of the GPD parameters (σ, ξ) is linked to the choice of the threshold u, which must be high enough for the asymptotic theory to apply, but not so high that too few exceedances lead to unstable estimates.

Naess (2024) notes that two common procedures are used for this purpose. The first is applied prior to model estimation and consists of examining whether the **mean residual life (MRL) plot** follows its theoretically expected pattern. Naess (2024) explains that this is based on the theoretical fact that, for a random variable X following a GPD exceeding a threshold u, the conditional mean excess function is linear in u. We next justify this property.

Let X be a random variable following a $GPD(\xi, \sigma)$, then, the expected value of X is:

$$\mathbb{E}[X] = \frac{\sigma}{1 - \xi}, \quad \forall \xi < 1 \tag{37}$$

Now, we assume that the GPD is a valid for the excesses above a fixed threshold u_0. Let Y denote an arbitrary observation among $Y_1, Y_2, \ldots, Y_n$. Then, using (37), we obtain:

$$\mathbb{E}[Y - u_0 \mid Y > u_0] = \frac{\sigma_{u_0}}{1 - \xi}, \tag{38}$$

with σ_u the scale parameter corresponding to the excesses above threshold u. If the GPD holds for threshold u_0, it must also hold for any $u > u_0$ and using the result in Appendix B:

$$\mathbb{E}[Y - u \mid Y > u] = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi(u - u_0)}{1 - \xi}, \tag{39}$$

Thus, for $u > u_0$, $\mathbb{E}[Y - u \mid Y > u]$ is a linear function of u. With n_u observations that exceed the threshold u and the excesses defined as $y_{(i)} - u$, the mean excess is then given by:

$$\hat{e}(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} \left(y_{(i)} - u\right), \tag{40}$$

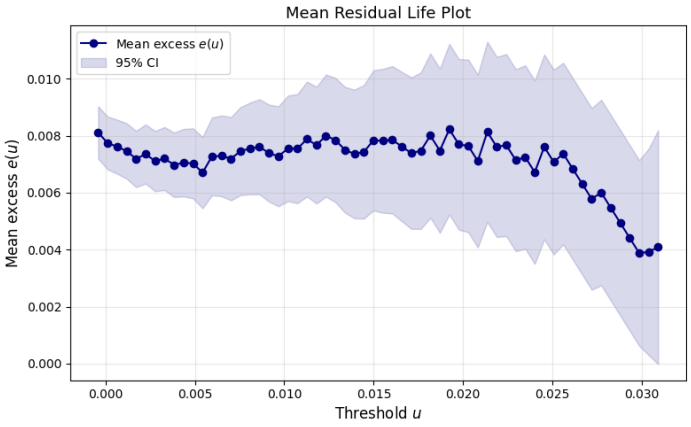
where $y_{(i)}$ denotes the ith ordered exceedance above u. Specifically, the points plotted in the empirical graph (MRL plot) are the following estimates:

$$\left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (y_{(i)} - u)\right), \quad u < y_{\max}, \tag{41}$$

with $y_{(1)}, y_{(2)}, \ldots, y_{(n_u)}$ the observations exceeding threshold u and $y_{\max}$ the largest observation.

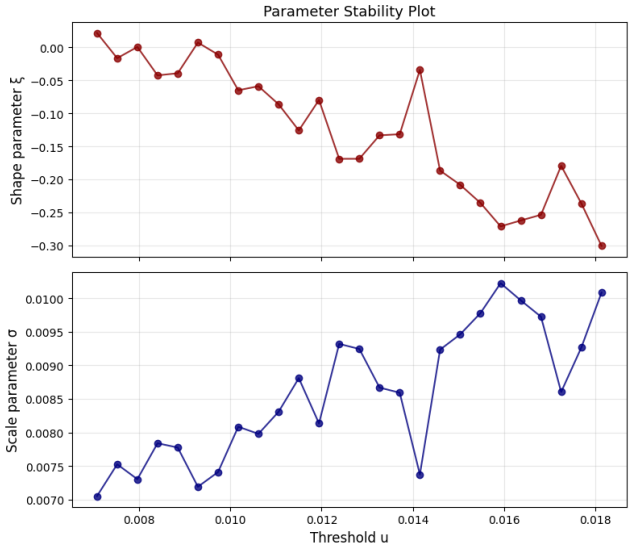
Figure 3 represents the MRL plot for our dataset. The approximately linear region in the MRL plot indicates the range where the GPD provides a suitable model for the excesses and the threshold u lies within this stable region. From plot, the mean excess function appears approximately stable for thresholds between $u \approx 0.007$ and $u \approx 0.02$.

Figure 3: Mean Residual Life plot for the dataset.



The second approach involves assessing the **stability of the parameter estimates** by fitting the GPD model over a range of candidate thresholds and identifying the region where the estimates remain relatively stable. Figure 4 displays the evolution of the estimated parameters as the threshold increases. The graphs show noticeable fluctuations across thresholds, without a clearly defined stable region, so there isn't an obvious choice of u.

Figure 4: Stability plot of the GPD parameter estimates for the dataset.



Although graphical diagnostics such as the MRL plot and the parameter stability plot provide valuable guidance for threshold selection, the results are not always conclusive. In our case, the visual evidence did not reveal a clearly linear or stable region from which to select u confidently. Therefore, we adopted a more pragmatic approach and set the threshold at a high quantile of the loss distribution, specifically the 90^{th} percentile, obtaining $u = 0.011$.

Appendix D: Construction of confidence intervals across VaR methods

In this appendix we summarise and formalise the construction of confidence intervals for the different Value-at-Risk estimation approaches employed in the empirical analysis.

D.1. Gaussian VaR

In the Gaussian framework, losses are assumed to follow a Gaussian distribution. In our application, the mean μ is treated as known (set to zero), while the variance σ^2 is unknown and must be estimated from the sample.

First, since $L_t \sim \mathcal{N}(\mu, \sigma^2)$ and μ is known, the standardised variables

$$Z_i = \frac{L_i - \mu}{\sigma}, \qquad i = 1, \dots, n, \tag{42}$$

are i.i.d. as $\mathcal{N}(0,1)$. As we know, the sum of squares of n standard normal variables satisfies $\sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$. Then, substituting Z_i we obtain:

$$\sum_{i=1}^{n} \left(\frac{L_i - \mu}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (L_i - \mu)^2 \sim \chi_n^2. \tag{43}$$

On the other hand, when the mean is known, the MLE of the variance is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (L_i - \mu)^2, \tag{44}$$

Then, the sampling distribution of the variance estimator satisfies

$$\frac{n\,\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (L_i - \mu)^2}{\sigma^2} \sim \chi_n^2. \tag{45}$$

From this representation, a $(1-\gamma)$ confidence interval for the true variance σ^2 is obtained by considering the central probability mass of the chi–square distribution:

$$\mathbb{P}\left(\chi^2_{n,1-\gamma/2} < \frac{n\,\hat{\sigma}^2}{\sigma^2} < \chi^2_{n,\gamma/2} \right) = 1 - \gamma. \tag{46}$$

Consequently, we obtain the confidence interval for the Gaussian VaR:

$$CI_{\text{VaR}}^{\text{Gaussian}} = \left[\mu + z_\alpha\, \sigma_{\text{lower}},\ \mu + z_\alpha\, \sigma_{\text{upper}} \right], \tag{47}$$

where:

$$\sigma_{\text{lower}} = \hat{\sigma} \sqrt{\frac{n}{\chi^2_{n,\,1-\gamma/2}}} \quad \text{and} \quad \sigma_{\text{upper}} = \hat{\sigma} \sqrt{\frac{n}{\chi^2_{n,\,\gamma/2}}}. \tag{48}$$

D.2. Nonparametric VaR

The historical estimator $\widehat{\text{VaR}}_\alpha(w; h)$ is defined from the empirical quantile $\widehat{F}_\Pi^{-1}(1-\alpha)$, its sampling distribution follows from classical order–statistic theory. Let $\Pi_{(1:n_S)}(w) \le \cdots \le \Pi_{(n_S:n_S)}(w)$ denote the ordered P&L values.

The probability integral transform if F_Π denotes the true P&L distribution, then $U_s = F_\Pi(\Pi_s(w)) \sim \text{Uniform}(0,1)$. Since F_Π is strictly increasing, ordering is preserved, and then we have $F_\Pi\big(\Pi_{(q:n_S)}(w)\big) = U_{(q)}$. Also, it is well known that the q-th order statistic of n_S i.i.d. uniform variables follows a Beta distribution, $U_{(q)} \sim \text{Beta}(q,\, n_S + 1 - q)$. Therefore,

$$F_\Pi\big(\Pi_{(q:n_S)}(w)\big) \sim \text{Beta}(q,\, n_S + 1 - q)\,, \tag{49}$$

a distribution-free identity valid for any continuous F_Π. Then, to construct a confidence interval for the quantile $F_\Pi^{-1}(1-\alpha)$, we set $q = \lfloor n_S(1-\alpha) \rfloor$. Then, the cumulative probability $p = F_\Pi\Big(\widehat{\text{VaR}}_\alpha(w;h)\Big)$ admits the exact confidence bounds

$$p_{\text{lower}} = \text{Beta}^{-1}\big(\tfrac{\gamma}{2};\, q,\, n_S + 1 - q\big)\,, \qquad p_{\text{upper}} = \text{Beta}^{-1}\big(1 - \tfrac{\gamma}{2};\, q,\, n_S + 1 - q\big)\,. \tag{50}$$

Mapping these probabilities back to the P&L domain using the empirical quantile function Q_{emp} yields the $(1-\gamma)$ confidence band:

$$\text{VaR}_{\alpha,\text{lower}} = Q_{\text{emp}}(p_{\text{lower}}), \qquad \text{VaR}_{\alpha,\text{upper}} = Q_{\text{emp}}(p_{\text{upper}}). \tag{51}$$

Thus, the nonparametric confidence interval for the Historical VaR is

$$CI_{\text{VaR}}^{\text{Hist}} = \big[\text{VaR}_{\alpha,\text{lower}},\ \text{VaR}_{\alpha,\text{upper}}\big]. \tag{52}$$

D.3. Extreme Value Theory

The bootstrap technique was introduced in 1979 as a computational method for estimating the standard error of $\hat{\theta}$. Bootstrap estimation of the standard error requires no theoretical derivation and can be applied regardless of how complicated the structure of $\hat{\theta} = s(x)$ may be. In this section we explain how this estimation is carried out.

Bootstrap methods rely on the notion of a *bootstrap sample*.

Definition 6.3 (Bootstrap sample)**.** Let $x = (x_1, \ldots, x_n)$ be a sample from an unknown distribution F, and let $\hat{F}$ denote its empirical distribution, which assigns probability $1/n$ to each observed value x_i. A *bootstrap sample* is an i.i.d. sample of size n drawn from $\hat{F}$:

$$x^* = (x_1^*, \ldots, x_n^*), \qquad x_j^* \sim \hat{F},\ j = 1, \ldots, n. \tag{53}$$

Therefore, we can generate bootstrap samples x^* by sampling with replacement from the original dataset of n observations, x, and compute the bootstrap replication of the estimator

$$\hat{\theta}^* = s(x^*). \tag{54}$$

where $s(x^*)$ is obtained by applying the same function $s(\cdot)$, originally applied to x, to x^*.

Let $\widehat{F}$ denote the empirical distribution function of the bootstrap replicates $\hat{\theta}^*$. The percentile bootstrap confidence interval of level $(1-\alpha)\cdot 100\%$ for θ is defined as the interval whose endpoints are the $\alpha/2$ and $(1-\alpha/2)$ empirical percentiles of $\widehat{F}$:

$$\big[\widehat{F}^{-1}(\alpha/2),\ \widehat{F}^{-1}(1-\alpha/2)\big]. \tag{55}$$

In practice, to compute this interval we proceed as follows.

1. We generate B bootstrap samples $x^{*,1}, \ldots, x^{*,B}$.
2. We compute the bootstrap replicate for each sample

$$\hat{\theta}^{*(b)} = s(x^{*,b}), \qquad b = 1, \ldots, B. \tag{56}$$

3. Construct the empirical distribution $\hat{\theta}^*_{B,\alpha}$ be the empirical α-quantile of the bootstrap replicates $\{\hat{\theta}^{*(b)}\}_{b=1}^B$.
4. Then, the percentile confidence interval of level $(1-\alpha)\cdot 100\%$ is given by

$$\left[\hat{\theta}^*_{B,\alpha/2},\ \hat{\theta}^*_{B,1-\alpha/2}\right]. \tag{57}$$

Bootstrap percentile intervals satisfy the following useful property: if $\varphi = m(\theta)$ is a monotone transformation of the parameter of interest, then the percentile confidence interval for φ is obtained simply by transforming the endpoints of the percentile interval for θ:

$$\left[m(\hat{\theta}^*_{B,\alpha/2}),\ m(\hat{\theta}^*_{B,1-\alpha/2})\right]. \tag{58}$$

This property is very useful in practice when the parameter of interest is a non-linear function of the data or model parameters, such as the VaR computed from GEV or GPD estimators.

D.3.1. Block maxima

In the Block Maxima framework, the parameter of interest is not one of the GEV parameters themselves, but the extreme quantile:

$$\theta = \mathrm{VaR}_\alpha^{\mathrm{GEV}} = F^{-1}_{\mathrm{GEV}}(\alpha_{\mathrm{GEV}}); \xi, \mu, \sigma) = m(\xi, \mu, \sigma), \tag{59}$$

which is a non–linear transformation of the GEV parameters (ξ, μ, σ). A *parametric bootstrap* is used, instead of sampling from the empirical distribution of the block maxima, we draw bootstrap samples from the fitted GEV distribution. The algorithm is the following.

1. Generate B bootstrap samples from the fitted GEV distribution:

$$x^{*,b} = (M_1^{*(b)}, \ldots, M_K^{*(b)}), \qquad M_i^{*(b)} \sim \mathrm{GEV}(\hat{\xi}, \hat{\mu}, \hat{\sigma}), \quad b = 1, \ldots, B. \tag{60}$$

2. For each bootstrap sample $x^{*,b}$, re-estimate the parameters via MLE.
3. Compute the bootstrap VaR replicate using the re-estimated parameters:

$$\hat{\theta}^{*(b)} = F^{-1}_{\mathrm{GEV}}\left(\alpha_{\mathrm{GEV}}; \hat{\xi}^{*(b)}, \hat{\mu}^{*(b)}, \hat{\sigma}^{*(b)}\right). \tag{61}$$

4. Let $\hat{\theta}^*_{B,\alpha}$ denote the empirical α–quantile of the bootstrap replicates $\{\hat{\theta}^{*(b)}\}_{b=1}^B$.
5. Then, by the percentile principle, the $(1-\alpha)\cdot 100\%$ confidence interval is:

$$CI^{\mathrm{GEV}}_{\mathrm{VaR}} = \left[\mathrm{VaR}^*_{\alpha,\mathrm{lower}},\ \mathrm{VaR}^*_{\alpha,\mathrm{upper}}\right] = \left[\hat{\theta}^*_{B,\alpha/2},\ \hat{\theta}^*_{B,1-\alpha/2}\right]. \tag{62}$$

D.3.2. POT

In the Peaks-Over-Threshold framework, exceedances above a high threshold u are modeled using the Generalized Pareto Distribution. The parameter of interest is again a high quantile of the original loss distribution, expressed as the POT VaR:

$$\theta = \mathrm{VaR}_\alpha^{\mathrm{POT}} = F^{-1}(\alpha) \;=\; u \;+\; \frac{\sigma}{\xi}\left[\left(\frac{n}{n_u}(1-\alpha)\right)^{-\xi} - 1\right], \tag{63}$$

which is a non–linear transformation of the GPD parameters (ξ, σ). A *parametric bootstrap* is employed: instead of resampling the observed exceedances, we draw bootstrap exceedance samples from the fitted GPD model. The algorithm is the following.

1. Generate B bootstrap samples from the fitted GPD distribution:

$$x^{*,b} = (Y_1^{*(b)}, \ldots, Y_{n_u}^{*(b)}), \qquad Y_i^{*(b)} \sim \mathrm{GPD}(\hat{\xi}, \hat{\tilde{\sigma}}), \quad b = 1, \ldots, B. \tag{64}$$

2. For each bootstrap exceedance sample $x^{*,b}$, re-estimate the parameters $(\xi, \tilde{\sigma})$ via MLE.
3. Compute the bootstrap POT VaR replicate using the re-estimated parameters:

$$\hat{\theta}^{*(b)} = u + \frac{\hat{\tilde{\sigma}}^{*(b)}}{\hat{\xi}^{*(b)}}\left[\left(\frac{p_{\mathrm{tail}}}{\zeta}\right)^{-\hat{\xi}^{*(b)}} - 1\right]. \tag{65}$$

4. Let $\hat{\theta}^*_{B,\alpha}$ denote the empirical α–quantile of the bootstrap replicates $\{\hat{\theta}^{*(b)}\}_{b=1}^B$.
5. Then, the percentile confidence interval of level $(1-\alpha)\cdot 100\%$ is:

$$CI_{\mathrm{VaR}}^{\mathrm{POT}} = \left[\mathrm{VaR}^*_{\alpha,\mathrm{lower}},\ \mathrm{VaR}^*_{\alpha,\mathrm{upper}}\right] = \left[\hat{\theta}^*_{B,\alpha/2},\ \hat{\theta}^*_{B,1-\alpha/2}\right]. \tag{66}$$

D.4. Time series

D.4.1. Filtered

Let $\{X_t\}_{t=1}^n$ denote the (nonstationary) return series and let $\widehat{\sigma}_t$ be a local estimate of the conditional volatility. The filtered series is defined as

$$Z_t = \frac{X_t}{\widehat{\sigma}_t}, \qquad t = 1, \ldots, n. \tag{67}$$

Under the assumption that $\{Z_t\}$ is approximately i.i.d., the VaR for day $T+1$ is:

$$\mathrm{VaR}_{T+1}(\alpha) = \widehat{z}_\alpha\, \widehat{\sigma}_{T+1}, \tag{68}$$

where $\widehat{z}_\alpha$ is the empirical α–quantile of the filtered sample $\{Z_t\}_{t=1}^n$ and $\widehat{\sigma}_{T+1}$ is the volatility forecast. Let $Z_{(1)} \le \cdots \le Z_{(n)}$ denote the order statistics of $\{Z_t\}$ and set $k = \lfloor \alpha n \rfloor$. The empirical quantile estimator is $\widehat{z}_\alpha = Z_{(k)}$.

A classical result on order statistics states that, for i.i.d. observations, we have that $F_Z\big(Z_{(k)}\big) \sim \mathrm{Beta}\big(k,\, n+1-k\big)$, where F_Z is the (unknown) distribution function of Z_t. Hence, a $(1-\alpha_{\mathrm{CI}})$ confidence interval for the cumulative probability level is given by:

$$p_{\mathrm{lower}} = B^{-1}\big(\alpha_{\mathrm{CI}}/2;\, k,\, n+1-k\big), \qquad p_{\mathrm{upper}} = B^{-1}\big(1-\alpha_{\mathrm{CI}}/2;\, k,\, n+1-k\big), \tag{69}$$

where $B^{-1}(\cdot;\, k,\, n+1-k)$ denotes the quantile function of the $\mathrm{Beta}(k, n+1-k)$ distribution.

Mapping these bounds back into quantile space using the empirical distribution of $\{Z_t\}$ yields

$$\widehat{z}_{\text{lower}} = \widehat{F}_Z^{-1}(p_{\text{lower}}), \qquad \widehat{z}_{\text{upper}} = \widehat{F}_Z^{-1}(p_{\text{upper}}), \tag{70}$$

where $\widehat{F}_Z$ is the empirical CDF of the filtered sample. Finally, the confidence interval for the filtered VaR on day $T+1$ is obtained by rescaling with the volatility forecast,

$$\big[\text{VaR}_{\alpha,\text{lower}}, \text{VaR}\,\alpha, \text{upper},\big] = \big[\widehat{z}_{\text{lower}}\,\widehat{\sigma}_{T+1}, \widehat{z}_{\text{upper}}\,\widehat{\sigma}_{T+1}\big]. \tag{71}$$

D.4.2. GARCH

The corresponding GARCH VaR at level α is:

$$\text{VaR}_\alpha^{\text{GARCH}} = \mu + z_\alpha\,\sigma_{T+1}, \tag{72}$$

where z_α denotes the standard normal quantile associated with the target tail probability. Under the Gaussian quasi-likelihood framework, the estimator satisfies the asymptotic chi–square identity as seen in C.1. Gaussian VaR confidence interval.

$$\frac{n\,\widehat{\sigma}_{T+1}^2}{\sigma_{T+1}^2} \sim \chi_n^2, \tag{73}$$

which yields an exact chi–square confidence interval for the conditional standard deviation:

$$\sigma_{\text{lower}} = \widehat{\sigma}_{T+1}\sqrt{\frac{n}{\chi^2_{n,\,1-\gamma/2}}}, \qquad \sigma_{\text{upper}} = \widehat{\sigma}_{T+1}\sqrt{\frac{n}{\chi^2_{n,\,\gamma/2}}}. \tag{74}$$

Thus, the confidence interval for the GARCH(1,1)-based VaR is

$$CI_{\text{VaR}}^{\text{GARCH}} = \big[\mu + z_\alpha\,\sigma_{\text{lower}},\ \mu + z_\alpha\,\sigma_{\text{upper}}\big]. \tag{75}$$

Appendix E: EVT model diagnostics

In this appendix we summarise the diagnostic tools used to evaluate the EVT models used. The aim is to assess whether the assumptions underlying the EVT limit results are reasonably satisfied and whether the fitted models provide a reliable description of the tail behaviour.

Block maxima approach

In order to assess whether the fitted GEV distribution $\widehat{G}$ provides an adequate description of the block maxima, M, we rely on a classical property of the GEV family. If $M \sim \text{GEV}(\mu, \sigma, \xi)$, then its cumulative distribution function is:

$$G_{\xi,\mu,\sigma}(M) = \exp\left\{-\left(1+\xi\left(\frac{M-\mu}{\sigma}\right)\right)^{-1/\xi}\right\} \qquad 1+\xi\left(\frac{M-\mu}{\sigma}\right) > 0. \tag{76}$$

Applying the Probability Integral Transform to M we know that $U = G_{\xi,\mu,\sigma}(M) \sim \text{Uniform}(0,1)$. Since U is uniform, then we see that $E \sim \text{Exp}(1)$:

$$E = -\log(U) = -\log\big(G_{\xi,\mu,\sigma}(M)\big) = \left(1+\xi\frac{M-\mu}{\sigma}\right)_+^{-1/\xi}, \tag{77}$$

Now, suppose that Y_i has distribution function F (unknown) and define the block maximum $M_n = \max(Y_1, \ldots, Y_n)$. The exact distribution of the block maximum is

$$P(M_n \leq y) = F(y)^n. \tag{78}$$

The Fisher–Tippett Theorem states that, there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that:

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} GEV, \qquad n \to \infty. \tag{79}$$

Consequently,

$$P(M_n \leq y) = P\left(\frac{M_n - b_n}{a_n} \leq \frac{y - b_n}{a_n}\right) \approx G\left(\frac{y - b_n}{a_n}\right). \tag{80}$$

Hence, to obtain the daily α–quantile of F, we solve $F(z_\alpha) = \widetilde{G}(z_\alpha)^{1/n} = 1 - \alpha$. Therefore,

$$z_\alpha = \widetilde{G}^{-1}((1-\alpha)^n), \tag{81}$$

which provides the daily VaR implied by the fitted GEV model.

Figure 5 displays the diagnostic plots used to assess whether the fitted GEV distribution provides an adequate description of the block maxima. All diagnostic plots rely on the fact that, under correct model specification, the transformed residuals should behave as i.i.d. $Exp(1)$. Equivalently, the empirical block maxima should follow the distribution implied by the fitted GEV model. Let us now analyse the four standard diagnostic plots shown in Figure 5.

Probability Plot. This plot compares the empirical cumulative probabilities of the block maxima with the theoretical probabilities computed from the fitted GEV distribution. If $M_{n,(i)}$ denotes the i-th order statistic of the block maxima and p_i the corresponding empirical plotting positions, then correct model specification implies $\widetilde{G}\big(M_{n,(i)}\big) \approx p_i$.

In the PP-plot, the empirical points remain reasonably close to the 45° reference line for intermediate probabilities, indicating that the fitted GEV captures the central part of the distribution well. However, a noticeable and systematic upward deviation appears as we move towards the upper tail: empirical probabilities exceed the theoretical ones. This suggests that the fitted GEV underestimates the probability of large block maxima.

Quantile Plot (QQ-plot). The QQ-plot compares the empirical block maxima with the theoretical quantiles of the fitted GEV distribution. Let $\widetilde{G}^{-1}$ denote the fitted quantile function, under the GEV model, $M_{n,(i)} \approx \widetilde{G}^{-1}(p_i)$.

The QQ-plot confirms the behaviour observed in the PP-plot. For small and moderate quantiles, the empirical maxima align well with the theoretical GEV quantiles. In contrast, the largest observations lie clearly above the diagonal line, indicating that the empirical tail is heavier than predicted by the model.

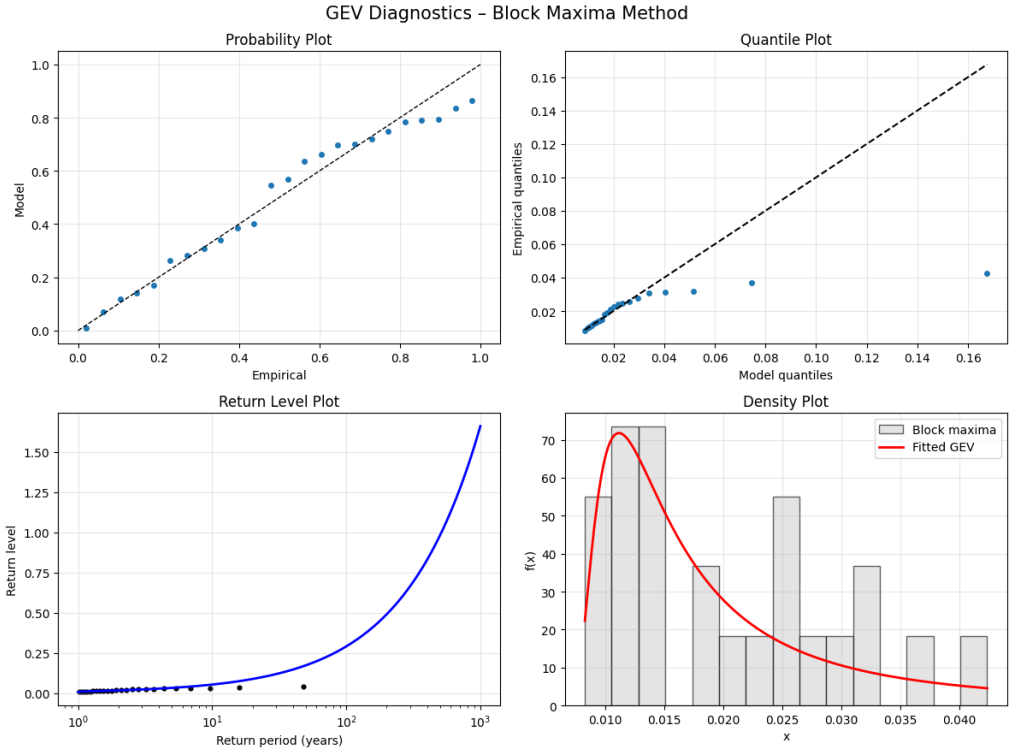
Return-Level Plot. The return-level plot displays the fitted return levels together with the empirical return levels computed from the ordered block maxima. The return level associated with a return period T is defined as the quantile z_T satisfying:

$$\mathbb{P}(M_n > z_T) = \frac{1}{T}, \quad \text{that is,} \qquad z_T = \widetilde{G}^{-1}\left(1 - \frac{1}{T}\right). \tag{82}$$

The return-level plot displays the same pattern: the empirical return levels follow the GEV curve for short return periods but fall consistently above the model as T increases. It implies that the fitted GEV distribution yields return levels that grow too slowly relative to the empirical data, reflecting underestimation of extreme quantiles.

Density Plot. The density plot compares the empirical distribution of block maxima with the fitted GEV density. In our case, the density plot shows that the fitted GEV density matches the bulk of the histogram reasonably.

Figure 5: Diagnostic plots for the GEV block maxima model.



POT approach

Under the Pickands–Balkema–de Haan theorem, for a sufficiently high u, the conditional excesses $Y_i = X_i - u \mid X_i > u$, should satisfy:

$$F_u(y) \approx G_{\xi,\sigma_u}(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-1/\xi}, \qquad y \geq 0, \tag{83}$$

where G_{ξ,σ_u} denotes the fitted GPD. Equivalently, the GPD-transformed residuals

$$E_i = -\log\big(1 - \widetilde{G}(Y_{(i)})\big) \tag{84}$$

should behave approximately as i.i.d. Exp(1) variables.

Figure 6 shows the diagnostic plots used to assess whether the fitted GPD adequately captures the behaviour of the exceedances above the selected threshold u.

Probability Plot. The PP-plot compares the empirical exceedance probabilities with the theoretical probabilities implied by the fitted GPD. If $Y_{(i)}$ denotes the i-th order statistic of the exceedances and p_i the corresponding empirical plotting positions, their expected plotting

positions, then correct model specification requires $\widetilde{G}(Y_{(i)}) \approx p_i$.

In the PP-plot, the empirical points track the diagonal very closely for most of the probability range, indicating that the fitted GPD captures the bulk of the exceedance distribution well. A mild deviation appears only near the upper tail, suggesting a slight underestimation of the most extreme exceedances.

Quantile Plot (QQ-plot). The QQ-plot compares the empirical exceedance quantiles with the theoretical GPD quantiles $\widetilde{G}^{-1}(p_i)$. Under the GPD model, $Y_{(i)} \approx \widetilde{G}^{-1}(p_i)$.

In the QQ-plot, empirical and model quantiles are almost indistinguishable over small and moderate ranges, showing an excellent agreement between the GPD and the data. Only a few of the largest exceedances fall slightly above the reference line, indicating a tail that is marginally heavier than the fitted GPD predicts.

Return-Level Plot. The return level associated with a return period T is defined as the level z_T satisfying

$$\mathbb{P}(X > z_T) = \frac{1}{T}. \tag{85}$$

Within the POT framework, exceedances above the threshold u are modelled by a GPD. Writing $\lambda = \mathbb{P}(X > u)$ for the exceedance probability, the tail probability for any $z > u$ is

$$\mathbb{P}(X > z) = \mathbb{P}(X > z \mid X > u)\, \mathbb{P}(X > u). \tag{86}$$

For $z > u$, the event $\{X > z\}$ is equivalent to $\{X - u > z - u\}$, and since we work conditionally on $X > u$, this becomes

$$\mathbb{P}(X > z \mid X > u) = \mathbb{P}(X - u > z - u \mid X > u) = \mathbb{P}(Y > z - u), \tag{87}$$

where $Y = X - u \mid X > u$ is the excess above the threshold u. Under the POT model, the excess $Y \sim G_{\xi,\sigma_u}(y)$, then:

$$\mathbb{P}(Y > y) = 1 - F_u(y) = \left(1 + \frac{\xi y}{\sigma_u}\right)^{-1/\xi}. \tag{88}$$

Therefore, we obtain the following:

$$\mathbb{P}(X > z) = \lambda \left(1 + \frac{\xi(z-u)}{\sigma_u}\right)^{-1/\xi}, \qquad z > u. \tag{89}$$

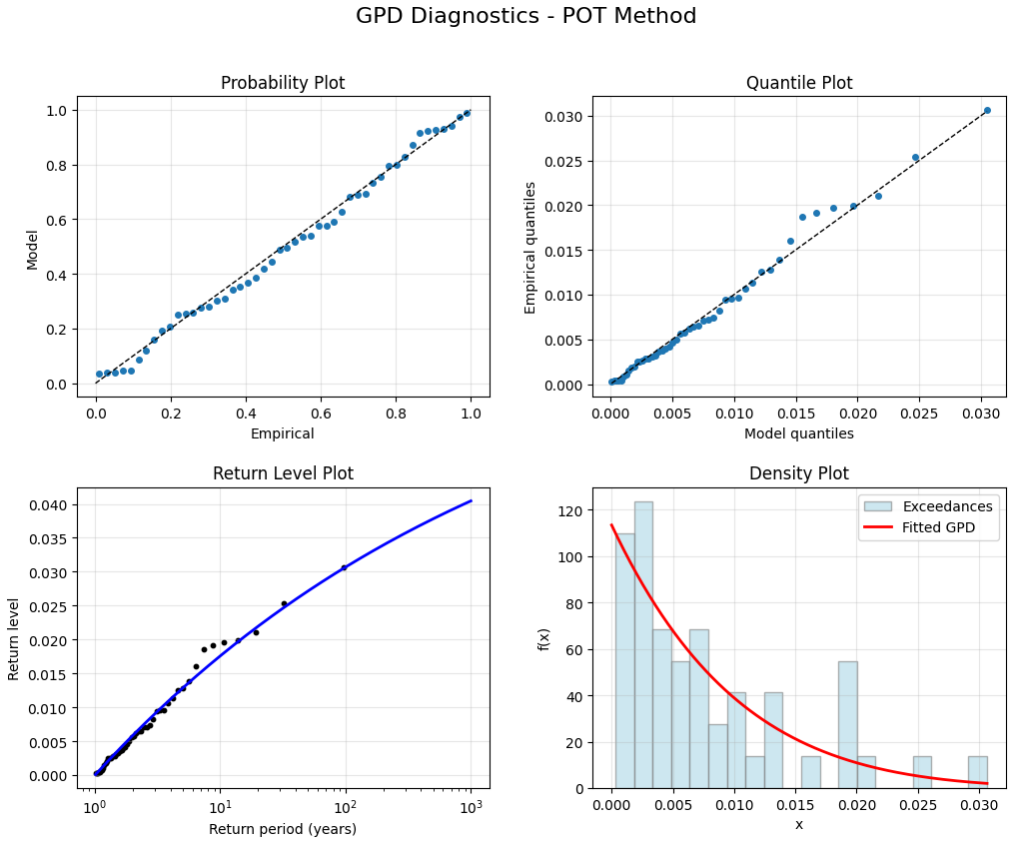
To obtain the return level z_T, we impose the defining condition (85), so we obtain:

$$\lambda \left(1 + \frac{\xi(z_T - u)}{\sigma_u}\right)^{-1/\xi} = \frac{1}{T} \iff z_T = u + \frac{\sigma_u}{\xi}\left((T\lambda)^{\xi} - 1\right). \tag{90}$$

The return-level plot shows very good agreement for short and medium return periods, where the empirical points lie close to the fitted curve. For large return periods, the empirical points fall slightly above the theoretical line, in line with the deviations observed in the PP- and QQ-plots and suggesting that the fitted GPD slightly underestimates extreme tail behaviour.

Density Plot. The density plot compares the empirical histogram of exceedances with the fitted GPD density. In our case, the fitted GPD density matches the overall shape of the histogram well.

Figure 6: Diagnostic plots for the GPD POT model.



Appendix F: GARCH(1,1) convergence to a continuous-time SV Model

In this appendix, we show that a GARCH(1,1) model can be seen as an approximation of a stochastic volatility model. Nelson (1990) was the first to demonstrate this connection. Nelson shows that when we observe data at very high frequency (that is, when the time between observations becomes very small), the discrete GARCH(1,1) equations start to behave like a continuous-time volatility model.

Let $X_t = \log S_t$ denote the log-price of an asset, where $(S_t)_{t>0}$ follows a Geometric Brownian Motion (GBM), given by:

$$dX_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma \, dW_t, \tag{91}$$

with W_t a standard Brownian motion. Then, assume we observe prices at instants $t_i = i/n$ for $i = 0, \ldots, n$, with n large. Then the log-return over a sampling interval is:

$$\varepsilon_{t_i} = X_{t_{i+1}} - X_{t_i} = \left(\mu - \frac{\sigma^2}{2}\right) \frac{1}{n} + \frac{\sigma}{\sqrt{n}} Z_{t_i}, \quad Z_{t_i} \sim \mathcal{N}(0,1). \tag{92}$$

Therefore, the log-return sequence (ε_{t_i}) is i.i.d. satisfying:

$$\varepsilon_{t_i} \sim \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right) \frac{1}{n}, \frac{\sigma^2}{n}\right). \tag{93}$$

This shows that, in a high-frequency setting, the mean of ε_{t_i} is of order $\mathcal{O}(1/n)$ and its standard deviation is of order $\mathcal{O}(1/\sqrt{n})$, with variance being constant. In particular, as $n \to \infty$, the

increments of the log-price become small. However, in real financial markets, volatility is not constant, so we assume that the log-returns follow a GARCH(1,1) model:

$$\begin{cases} \varepsilon_{t_i} = \dfrac{\sigma_{t_i}}{\sqrt{n}} Z_{t_i}, \\ \sigma_{t_i}^2 = \omega_n + \alpha_n \varepsilon_{t_{i-1}}^2 + \beta_n \sigma_{t_{i-1}}^2, \end{cases} \tag{94}$$

Then, substituting $\varepsilon_{t_{i-1}} = \dfrac{\sigma_{t_{i-1}}}{\sqrt{n}} Z_{t_{i-1}}$ into the volatility equation:

$$\sigma_{t_i}^2 = \omega_n + \left(\frac{\alpha_n}{n} + \beta_n \right) \sigma_{t_{i-1}}^2 + \frac{\alpha_n}{n} \sigma_{t_{i-1}}^2 (Z_{t_{i-1}}^2 - 1). \tag{95}$$

Hence, by subtracting $\sigma_{t_{i-1}}^2$ from both sides of the previous expression, we obtain the incremental form of the conditional variance:

$$\sigma_{t_i}^2 - \sigma_{t_{i-1}}^2 = \omega_n - \left(1 - \frac{\alpha_n}{n} - \beta_n \right) \sigma_{t_{i-1}}^2 + \frac{\alpha_n}{n} \sigma_{t_{i-1}}^2 (Z_{t_{i-1}}^2 - 1), \tag{96}$$

where $\sigma_{t_{i-1}}^2 (Z_{t_{i-1}}^2 - 1)$ is a centered noise term (i.e. has mean zero). In order for the discrete GARCH recursion to converge to a stochastic differential equation, the size of its increments must match the characteristic scales of an SDE. Then, the parameters are scaled so that the drift terms are of order $\mathcal{O}(1/n)$ and the stochastic component is of order $\mathcal{O}(1/\sqrt{n})$:

$$\begin{cases} \dfrac{\alpha_n}{n} = \dfrac{\alpha_0}{\sqrt{n}} \iff \alpha_n = \sqrt{n}\, \alpha_0, \\ \omega_n = \dfrac{\omega_0}{n}, \\ 1 - \dfrac{\alpha_n}{n} - \beta_n = \dfrac{\beta_0}{n} \iff \beta_n = 1 - \dfrac{\beta_0}{n} - \alpha_0 \sqrt{n}. \end{cases} \tag{97}$$

The scaling relations in (105) are the key step that links the discrete GARCH(1,1) recursion to the continuous-time SV diffusion. These conditions are imposed so that each component of the GARCH increment in (96) has the correct asymptotic order as the sampling frequency n increases (i.e., as the time step $dt = 1/n$ tends to zero). Thus, when the sampling interval becomes small ($n \to \infty$), the GARCH parameters behave as:

$$\begin{cases} \omega_n \approx \omega_0\, dt, \quad \omega_n \to 0, \\ \alpha_n \approx \alpha_0\, dt, \quad \alpha_n \to 0, \\ 1 - \beta_n \approx \beta_0\, dt, \quad \beta_n \to 1, \\ \alpha_n + \beta_n \to 1. \end{cases} \tag{98}$$

Therefore, for $Z_{t_{i-1}} \sim$ i.i.d. $\mathcal{N}(0,1)$, we obtain the following increment representation:

$$\begin{cases} \varepsilon_{t_i} = X_{t_i} - X_{t_{i-1}} = \dfrac{\sigma_{t_{i-1}}}{\sqrt{n}}\, Z_{t_{i-1}}, \\ \sigma_{t_i}^2 - \sigma_{t_{i-1}}^2 = \dfrac{\omega_0}{n} - \dfrac{\beta_0}{n}\, \sigma_{t_{i-1}}^2 + \dfrac{\alpha_0}{\sqrt{n}}\, \sigma_{t_{i-1}}^2\, (Z_{t_{i-1}}^2 - 1), \end{cases} \tag{99}$$

The first term in (99) is a deterministic drift term of order $\mathcal{O}(1/n)$, while the last term contains a centered random component of order $\mathcal{O}(1/\sqrt{n})$. To characterize the limiting behaviour of this stochastic noise term when $n \to \infty$, we apply the Multivariate Central Limit Theorem.

Theorem 6.1 (Multivariate Central Limit Theorem)**.** *Let $\{Y_i\}_{i\geq 1}$ be i.i.d. random vectors in $\mathbb{R}^d$ with $\mathbb{E}[Y_1] = \mu \in \mathbb{R}^d$ and $\mathrm{Cov}(Y_1) = \Sigma \in \mathbb{R}^{d\times d}$, where Σ is positive definite. Then,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \mu) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d(0, \Sigma). \tag{100}$$

In our case, we define the vector of standardized innovations as:

$$Y_i = \begin{pmatrix} Z_{t_i} \\ Z_{t_i}^2 - 1 \end{pmatrix}, \qquad Z_{t_i} \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

Since $\mathbb{E}[Z_{t_i}] = 0$ and $\mathbb{E}[Z_{t_i}^2 - 1] = 0$, we deduce that:

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \tag{101}$$

Therefore, by applying the Multivariate Central Limit Theorem, we obtain the following:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} Z_{t_i} \\ Z_{t_i}^2 - 1 \end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right). \tag{102}$$

Finally, to connect the discrete and continuous formulations, we interpret the normalized discrete innovations as stochastic differentials:

$$\frac{Z_{t_{i-1}}}{\sqrt{n}} \approx dW_t^{(1)}, \qquad \frac{Z_{t_{i-1}}^2 - 1}{\sqrt{n}} \approx \sqrt{2}\, dW_t^{(2)}. \tag{103}$$

As $n \to \infty$, the pair $(X_{t_i}, \sigma_{t_i}^2)$ converges in law to the solution of the following continuous-time SV system:

$$\begin{cases} dX_t = \sigma_t\, dW_t^{(1)}, \\ d\sigma_t^2 = (\omega_0 - \beta_0 \sigma_t^2)\, dt + \sqrt{2\alpha_0}\, \sigma_t^2\, dW_t^{(2)}, \end{cases} \tag{104}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are independent standard Brownian motions.

Appendix G: Simulation Framework for GARCH–SV Convergence

In this appendix, we investigate the theoretical link between discrete-time GARCH models and continuous-time stochastic volatility diffusions. To do so, we simulate high-frequency prices from a stochastic volatility SDE, estimate GARCH(1,1) models directly from the simulated returns, and examine how the estimated parameters evolve under different sampling frequencies. This provides an empirical illustration of the well-known result that, under temporal aggregation, GARCH(1,1) parameters converge toward the continuous-time stochastic volatility limit, in particular through the behaviour of the persistence parameter $\alpha + \beta$. After analysing the simulated dataset, we repeat the same procedure on real tick-by-tick price data to assess whether the convergence patterns observed in controlled experiments remain valid in practice.

G.1. Simulated data

G.1.1. Stochastic Volatility simulation

We simulate the log-price X_t and its instantaneous variance σ_t^2 according to the stochastic volatility diffusion expressed in (26). This continuous-time specification provides the dynamics from which we later extract discrete-time returns and analyse how well a GARCH(1,1) model recovers the underlying volatility behaviour.

First, the GARCH parameters are rescaled as functions of the sampling frequency, as explained in Appendix F.

$$\begin{cases} \dfrac{\alpha_n}{n} = \dfrac{\alpha_0}{\sqrt{n}} \iff \alpha_n = \sqrt{n}\,\alpha_0, \\ \omega_n = \dfrac{\omega_0}{n}, \\ 1 - \dfrac{\alpha_n}{n} - \beta_n = \dfrac{\beta_0}{n} \iff \beta_n = 1 - \dfrac{\beta_0}{n} - \alpha_0\sqrt{n}. \end{cases} \tag{105}$$

Under this scaling, the discrete-time GARCH variance dynamics converge to the stochastic volatility diffusion:

$$d\sigma_t^2 = (\omega_0 - \beta_0\sigma_t^2)\,dt + \sqrt{2\alpha_0}\,\sigma_t^2\,dW_t, \tag{106}$$

We simulate $D = 10$ trading days, each containing $n = 10{,}000$ ticks, for a total of 100,000 observations of the log-price process. The resulting dataset captures rapid fluctuations in volatility, mimicking realistic intraday dynamics.

Later, to generate high-frequency trajectories of the log-price and its instantaneous variance, the continuous-time stochastic volatility system is discretized using an **Euler–Maruyama scheme** on a fine time grid.

$$\begin{cases} X_{k+1} &= X_k + \sigma_k\sqrt{\Delta t}\,Z_k^{(1)}, \\ \sigma_{k+1}^2 &= \sigma_k^2 + (\omega_0 - \beta_0\sigma_k^2)\,\Delta t + \sqrt{2\alpha_0}\,\sigma_k^2\sqrt{\Delta t}\,Z_k^{(2)}, \end{cases} \tag{107}$$

where Δt denotes the time step and $(Z_k^{(1)}, Z_k^{(2)})$ are i.i.d. standard normal random variables $Z_k^{(1)}, Z_k^{(2)} \sim \mathcal{N}(0,1)$, and, unless stated otherwise, they are taken to be independent for all k. This scheme produces a high-frequency trajectory $\{(X_k, \sigma_k^2)\}_{k=0}^N$ for both the log-price and the instantaneous variance.

Finally, we obtain discrete-time returns by differencing:

$$\epsilon_i = X_i - X_{i-1}, \qquad i = 1, \ldots, n-1. \tag{108}$$

These returns serve as the input for the subsequent GARCH(1,1) estimation, allowing us to test how well the discrete-time model approximates the continuous-time SV structure.

G.1.2. GARCH(1,1) estimation

The discrete-time GARCH(1,1) model is defined by

$$\sigma_i^2 = \omega + \alpha\,\epsilon_{i-1}^2 + \beta\,\sigma_{i-1}^2, \quad i = 1, \ldots, n-1, \tag{109}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$ for stationarity. We estimate (ω, α, β) using maximum likelihood estimation under the assumption that $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$.

Subsampling and Parameter convergence

To study the effect of sampling frequency on the GARCH parameters, we subsample the high-frequency returns at multiple rates $m \in \{1, 5, 10, 50, 100\}$, obtaining the subsampled series

$$\epsilon_i^{(m)} = \epsilon_{1+(i-1)m}, \quad i = 1, \ldots, \left\lfloor \frac{n-1}{m} \right\rfloor. \tag{110}$$

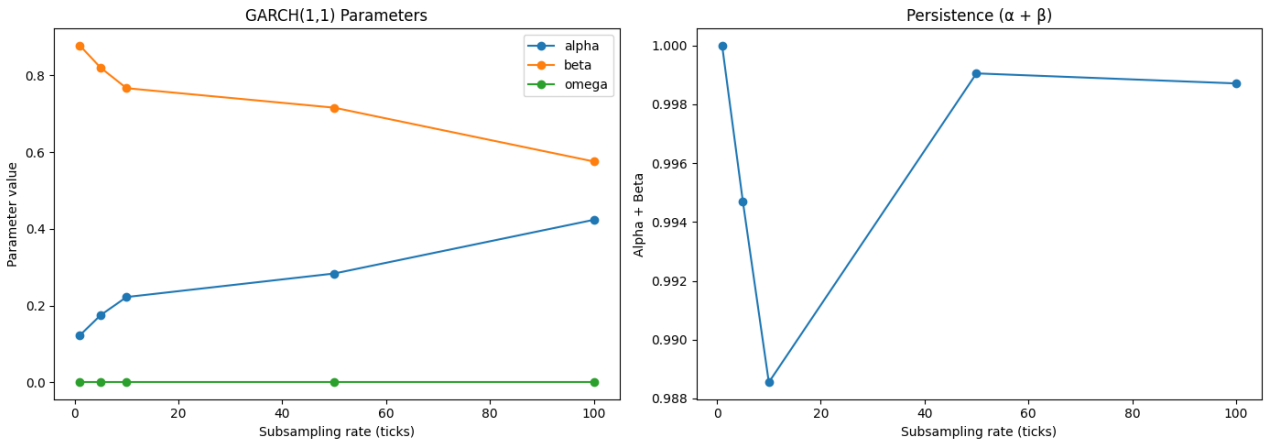
For each subsampled series, we refit the GARCH(1,1) model and record the estimated parameters $(\omega^{(m)}, \alpha^{(m)}, \beta^{(m)})$. This allows us to observe how the estimated parameters vary as a function of the sampling frequency. In particular, theory predicts that as the sampling frequency decreases, $\alpha^{(m)} + \beta^{(m)} \to 1$, reflecting the persistence of variance in the continuous-time stochastic volatility limit.

G.1.3. Results

We first plot a sample of the simulated log-price series X_t and the true variance σ_t^2 to illustrate the stochastic volatility dynamics. Then, we compare the GARCH(1,1) estimated conditional variance $\hat{\sigma}_i^2$ with the true variance. We also examine the contributions of the $\alpha\epsilon_{i-1}^2$ and $\beta\sigma_{i-1}^2$ terms to the conditional variance over time. Finally, we plot the estimated GARCH parameters $(\omega^{(m)}, \alpha^{(m)}, \beta^{(m)})$ as a function of the subsampling rate m to study their convergence, as well as the sum $\alpha^{(m)} + \beta^{(m)}$.

Figure 7 presents the convergence of the GARCH(1,1) parameter estimates and the persistence measure as the sampling frequency varies in the simulated data. In Figure 7, the right panel shows the persistence measure $\alpha + \beta$. For all subsampling levels, this quantity remains very close to one. This pattern reflects the fact that, when volatility is driven by an underlying continuous-time stochastic process, a discrete-time GARCH model fitted at lower frequencies must exhibit high persistence in order to capture the slow evolution of volatility. Although the individual estimates of α and β change with the sampling frequency, their sum behaves as predicted by theory, converging towards one. This result supports the interpretation of the GARCH(1,1) model as a discrete-time approximation of the continuous-time volatility dynamics.

Figure 7: Convergence of GARCH(1,1) parameter estimates and persistence as a function of sampling frequency as a function of sampling frequency for simulated data.



G.2. Application to real High-Frequency data

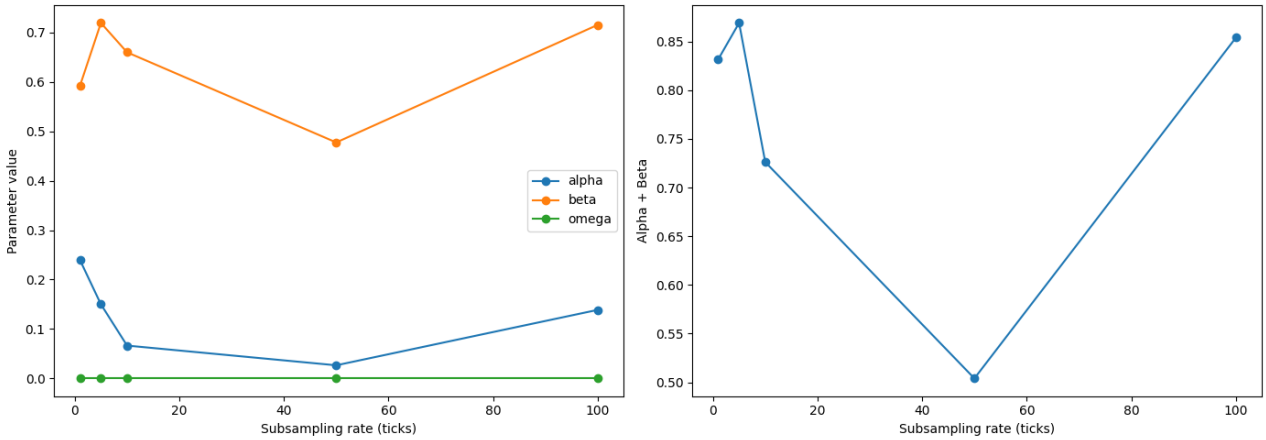
After gaining insights from the simulated data, we applied the same procedure to a real high-frequency dataset. The data consists of tick-by-tick observations with three columns: `day`, `tick`, and `logprice`. We first computed log-returns:

$$\epsilon_i = X_i - X_{i-1}, \quad X_i = \text{logprice at tick } i \tag{111}$$

and then fitted a **GARCH(1,1)** model from scratch using maximum likelihood estimation. We subsampled the returns at rates of 1, 5, 10, 50, and 100 ticks, and fitted the GARCH model on each subsampled series to study the effect of sampling frequency. This allowed us to observe the convergence behavior of the parameters ω, α, and β as a function of the sampling interval.

Figure 8 presents the convergence of the GARCH(1,1) parameter estimates and the persistence measure as the sampling frequency varies in the real data. The right panel reports the persistence measure $\alpha + \beta$. Unlike the simulated case, persistence is not consistently close to one. This behaviour highlights the impact of market microstructure effects, estimation noise, and model misspecification when applying GARCH models to real high-frequency data.

Figure 8: Convergence of GARCH(1,1) parameter estimates and persistence as a function of sampling frequency for real high-frequency data.



G.3. Computation of VaR with high frequency time series data

Using the GARCH(1,1) specification introduced in 2.4.2, we adapt the model to the high-frequency setting. Tick-by-tick returns are typically very small, exhibit strong volatility clustering, and have negligible conditional mean. For this reason, once the GARCH parameters have been estimated from high-frequency data, the one-step-ahead volatility forecast obtained from the model can be directly used to compute intraday risk measures.

Let (X_t) denote the high-frequency log-returns. Since $\mu_t \approx 0$ at very short horizons, the GARCH(1,1) forecast σ_{t+1} derived earlier simplifies the VaR computation. In particular, because the next-period innovation satisfies: $\varepsilon_{t+1} = \sigma_{t+1} Z_{t+1}$, where $Z_{t+1} \sim \mathcal{N}(0,1)$, the conditional α-quantile of X_{t+1} reduces to a rescaling of the normal quantile:

$$\text{VaR}_X(\alpha)_{t+1} = \sigma_{t+1}\, z_\alpha. \tag{112}$$

This formulation yields a fully dynamic intraday VaR series: as each new high-frequency return

arrives, the GARCH recursion updates σ_{t+1} at each tick, and the VaR adjusts immediately to reflect rapid changes in volatility, driven by market microstructure effects.

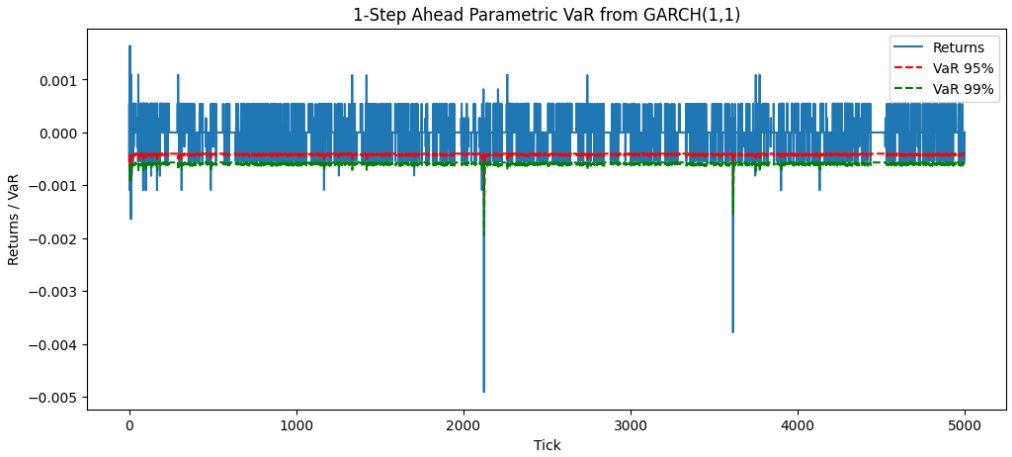
After fitting the GARCH(1,1) model to the full high-frequency return series, we computed the one-step-ahead VaR using the estimated parameters $\widehat{\omega} = 0.00000$, $\widehat{\alpha} = 0.02642$, and $\widehat{\beta} = 0.47551$ (obtained from the $m = 1$ tick-by-tick series).

The first five one-step-ahead VaR values obtained are:

$$\text{VaR}_{1:5} = (-0.00060251,\ -0.00062122,\ -0.00062992,\ -0.00063401,\ -0.00073043).$$

Figure 9 displays the full series of one-step-ahead VaR alongside the high-frequency returns. The plot clearly illustrates how the dynamic VaR bounds (both positive and negative) track the instantaneous volatility of the tick-by-tick returns, successfully capturing periods of increased market activity and risk.

Figure 9: One-step-ahead VaR using GARCH(1,1) on high-frequency returns.



G.4. Generalized GARCH(p,q) and Model Selection

To ensure the GARCH(1,1) specification is the most appropriate for our high-frequency dataset, we implement a generalized GARCH(p, q) framework. This allows for higher-order dependencies in both the shock (ARCH) and persistence (GARCH) components. The conditional variance for the GARCH(p, q) model is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{113}$$

We utilize three primary information criteria to penalize model complexity and prevent overfitting. Let L be the maximum likelihood, k the number of estimated parameters $(1+p+q)$, and n the number of observations:

- **AIC (Akaike Information Criterion):** $2k - 2\ln(L)$
- **BIC (Bayesian Information Criterion):** $k\ln(n) - 2\ln(L)$
- **AICC (Corrected AIC):** $AIC + \frac{2k^2+2k}{n-k-1}$

We tested three specifications: GARCH(1,1), GARCH(1,2), and GARCH(2,1), on a dataset of $n = 1,721,979$ tick observations. The results of the Maximum Likelihood Estimation and the corresponding information criteria are summarized in Table 5.

Table 5: Model Selection Summary for GARCH(p, q) Specifications.

Model	k	**Log-L**	**AIC**	**BIC**	**AICC**
GARCH(1,1)	3	1.1829×10^7	-2.3657×10^7	-2.3657×10^7	-2.3657×10^7
GARCH(1,2)	4	1.1792×10^7	-2.3584×10^7	-2.3584×10^7	-2.3584×10^7
GARCH(2,1)	4	-6.4516×10^6	1.2903×10^7	1.2903×10^7	1.2903×10^7

The estimation yielded the following parameters for the leading model, GARCH(1,1):

- $\hat{\omega} = 0.000000$
- $\hat{\alpha}_1 = 0.465494$
- $\hat{\beta}_1 = 0.452962$

As shown in Table 5, the **GARCH(1,1)** model achieves the highest Log-Likelihood and the lowest scores across all information criteria (AIC, BIC, and AICC). The GARCH(2,1) model performed significantly worse, indicated by a negative Log-Likelihood and positive information criteria scores. Given that both AIC and BIC converge on the same result, we conclude that the **GARCH(1,1)** is the optimal model for this high-frequency dataset. This justifies our use of the (1,1) specification for the dynamic VaR calculations performed in section G.3.