

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



**Modelado de rendimientos financieros:
volatilidad condicional, colas pesadas y
dependencia multivariada**

TESIS

QUE PARA OBTENER EL TÍTULO DE
**LICENCIADO EN ACTUARÍA Y DIRECCIÓN
FINANCIERA**

PRESENTA

FERNANDO PÉREZ MILLÁN

ASESOR

**DR. CARLOS VLADIMIR RODRÍGUEZ
CABALLERO**

«Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada **“Confianza institucional, inclusión social, intensidad religiosa y percepción tecnológica como factores de innovación para el crecimiento económico”**, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación.»

FECHA

FERNANDO PÉREZ MILLÁN

A Fátima.

Contents

Introduction	1
1 Empirical Characteristics of Returns	3
2 Volatility Modeling	10
3 Shock Modeling	18
4 Forecasting and Evaluation	36
5 Copula Multivariate Model	61
Conclusion	78
References	81

Introduction

A PhD student sits in his advisor's office, scrambling through piles of computer outputs in an attempt to finish a task his advisor assigned. A visiting professor enters, and their ensuing conversation leads to a question so intriguing that it consumes the student's thoughts overnight. The next morning, he rushes to the lab and begins coding. This is the origin story of the GARCH model, as narrated by [Bollerslev \(2023\)](#), who describes the creation of the model as a personal odyssey. At that moment, Bollerslev had little idea of the profound impact his model would have, revitalizing and significantly expanding the literature on financial return modeling. This literature quickly grew, matured, and branched out, generating numerous alternative approaches for nearly every aspect of return modeling. This thesis is, in essence, a consequence of the question posed in 1984 and the extensive literature it subsequently inspired.

The goal of this thesis is to provide a comprehensive overview of financial return modeling. While the thesis acknowledges and discusses alternative methodologies at various modeling stages, it follows a specific path designed to provide the reader with a clear, coherent, and complete

narrative. Ultimately, it aims to serve as a concise yet thorough "recipe" for modeling financial returns at the undergraduate level.

Chapter 1 outlines the fundamental task of return modeling, highlighting common empirical characteristics that models must capture and replicate. Chapters 2 through 4 then guide the reader through the univariate modeling process. Chapter 2 addresses volatility modeling, Chapter 3 focuses on modeling the distribution of shocks, and Chapter 4 presents methods for rigorously evaluating and testing these models. Chapter 5 extends the framework to a multivariate setting, building coherently on the foundations established in the previous chapters.

All  code used to generate the analyses and figures is available in the accompanying GitHub repository: github.com/ferpmillan/thesis.

Chapter 1

Empirical Characteristics of Returns

The modeling of asset returns has numerous applications, ranging from risk management to asset valuation over time. It is built upon a multi-theoretical foundation in time series analysis and financial theory, drawing insights from a wide range of disciplines. However, modeling asset returns remains an empirical craft, and practical relevance is demanded of these models. This should not come as a surprise, for financial markets are not mere theoretical abstractions; they are living, breathing systems that evolve and play a crucial role in the global economy.

Campbell, Lo, and MacKinlay (1997) give two main reasons for using returns instead of prices. First, for the average investor, the return of an asset provides a complete and scale-free summary of the investment opportunity. Second, for theoretical and empirical reasons that will become apparent, return series have attractive statistical properties.

We start by defining the daily continuously compounded or log return of an asset from the closing prices of the asset:

$$R_{t+1} = \ln(S_{t+1}) - \ln(S_t)$$

where $\ln(*)$ denotes the natural logarithm.

Most financial asset returns exhibit common empirical characteristics—what [Christoffersen \(2012\)](#) calls the stylized facts of asset returns and what [Tsay \(2010\)](#) refers to as the empirical properties of returns. **These empirical characteristics serve as guidelines for what models should capture and replicate.** To illustrate each of these characteristics, we will use daily log returns of the S&P 500 from January 1, 2014, to December 31, 2024.

Daily log returns exhibit very little autocorrelation. Formally, we can write

$$\text{Corr}(R_{t+1}, R_{t+1-\tau}) \approx 0 \quad \text{for } \tau = 1, 2, 3, \dots, 100.$$

This implies that past returns possess little to no predictive power for future returns. Figure 1.1 shows the correlation of daily S&P 500 log returns with returns lagged from 1 to 100 days.

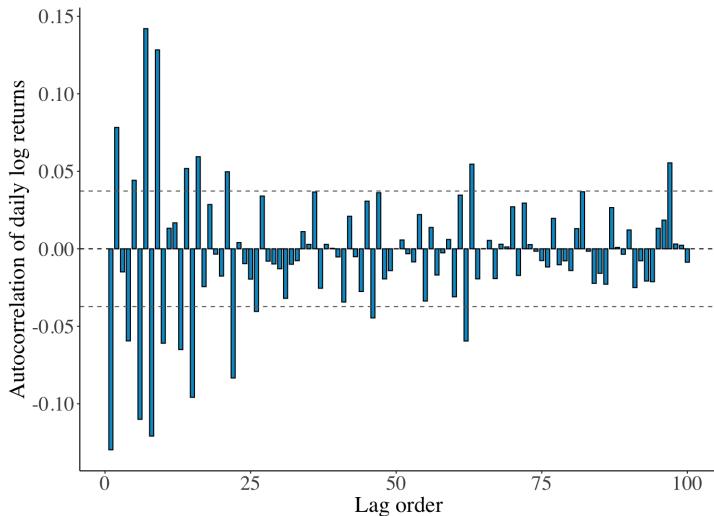


Figure 1.1: Autocorrelation of daily S&P 500 log returns from January 1, 2014, to December 31, 2024.

The unconditional distribution of daily returns departs from normality. Figure 1.2 shows a histogram of S&P 500 daily log returns together with a normal density fitted by maximum likelihood. Several features stand out. First, the empirical distribution is more peaked around zero (i.e., leptokurtic). Second, it exhibits heavier tails than the normal, implying a higher probability of extreme outcomes. Third, the distribution is asymmetric: in this sample it displays left-skewness (negative skewness), with a longer left tail than the right. Accurately modeling tail behavior is crucial for risk management, as emphasized by Christoffersen (2012).

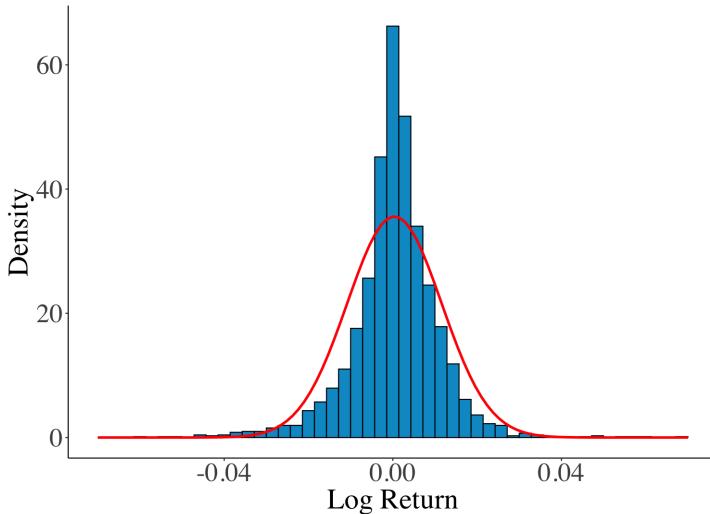


Figure 1.2: Histogram of daily S&P 500 log returns and the normal distribution January 1, 2014 - December 31, 2024. Skewness : -0.8088,
Kurtosis : 19.0445

The stock market experiences sharp declines in stock prices and index values, but not equally large upward movements. Consequently, the return distribution exhibits negative skewness.

The standard deviation of returns is considerably larger than the mean at short horizons, such as daily. Statistically, it is reasonable not to reject a zero mean return. Our S&P 500 data exhibit a daily mean return of 0.0380% and a daily standard deviation of 1.1211%.

Volatility measures, such as squared returns, exhibit positive correlation with their past values, reflecting the tendency of high-volatility events to cluster over time. This effect is most pronounced at short horizons,

such as daily returns. Formally, we can write

$$\text{Corr}(R_{t+1}^2, R_{t+1-\tau}^2) > 0 \quad \text{for small } \tau.$$

Figure 1.3 shows the autocorrelation in squared returns for the S&P 500 data. Models that account for this dependence will be introduced in the next section.

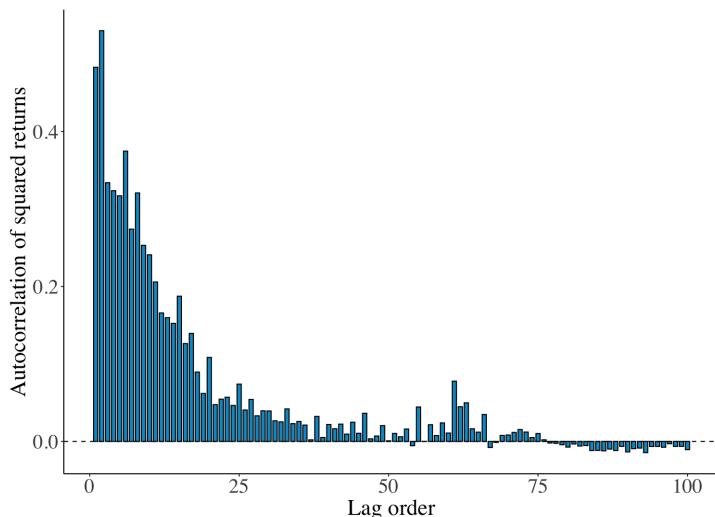


Figure 1.3: Autocorrelation of squared daily S&P 500 log returns from January 1, 2014, to December 31, 2024.

For equities and equity indices, asset volatility is negatively correlated with returns. A negative return increases the (conditional) variance more than a positive return of the same magnitude. This phenomenon is known as the leverage effect, arising from the fact that a decrease in a stock's price increases the firm's leverage—assuming its debt remains constant—thereby making it riskier.

The correlation among assets appears to vary over time. Notably, it tends to increase in magnitude during highly volatile bear markets and spikes dramatically during market crashes. Figure 1.4 displays the rolling covariance between S&P 500 returns and the returns on a 10-year Treasury note index, calculated using a 25-day rolling window. A sharp increase in covariance is evident during the early phase of the 2020 pandemic.

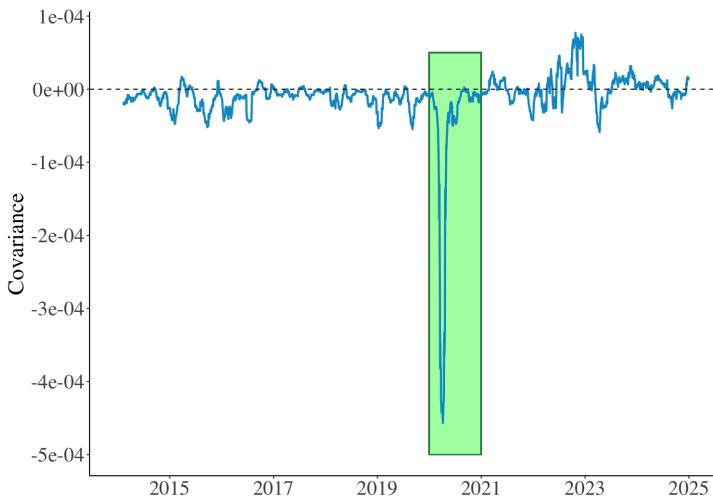


Figure 1.4: Rolling covariance between S&P 500 and 10-year treasury note index.

Lastly, even after standardizing returns by time-varying volatility, they continue to exhibit heavier tails than the normal distribution. However, the tails are lighter than those in the unconditional return distribution. This provides evidence of conditional non-normality.

[Cont \(2001\)](#) provides a detailed discussion of the empirical characteristics of returns presented here, along with additional ones, and examines various statistical challenges that arise in the analysis of financial time series.

Considering the empirical behavior of returns, the model for individual asset returns is

$$R_{t+1} = \mu_{t+1} + \sigma_{t+1} z_{t+1} \quad \text{where } z_{t+1} \sim \text{i.i.d. } D(0, 1). \quad (1.1)$$

The random variable z_{t+1} represents the innovation term or shock, which we assume to be identically and independently distributed (i.i.d.) with zero mean and unit variance. We assume that the conditional mean of returns, μ_{t+1} , is zero — a reasonable assumption based on the empirical behavior of daily returns discussed earlier. For longer horizons a model for the mean can be necessary¹, here we focus on modeling daily returns.

The specification for daily returns is then given by

$$R_{t+1} = \sigma_{t+1} z_{t+1} \quad \text{where } z_{t+1} \sim \text{i.i.d. } D(0, 1). \quad (1.2)$$

The univariate and multivariate return models developed in the subsequent chapters are designed to capture and replicate the empirical properties documented in this section. The next section models the conditional volatility, σ_{t+1} , and the following section models the standardized innovation, z_{t+1} .

¹ Readers interested in modeling the conditional mean, μ_{t+1} , are referred to [Tsay \(2010\)](#) Chapters 2 and 4, and the references therein. [Tsay \(2010\)](#) provides an exhaustive review of linear and nonlinear models for the conditional mean in financial time series.

Chapter 2

Volatility Modeling

The objective of this chapter is to introduce econometric methods commonly used to model the conditional volatility of asset returns, σ_{t+1} . These models are collectively referred to as conditional heteroscedastic models.

As noted by [Tsay \(2010\)](#), a distinctive feature of stock volatility is that it is not directly observable. Daily volatility cannot be directly inferred from return data because there is only one observation per trading day. However, despite its unobservability, volatility exhibits certain characteristics that are consistently observed in asset returns.

First, as established in the previous chapter, volatility—measured by squared returns—exhibits strong autocorrelation. This implies that if a recent period experienced high variance, the following day is also likely to exhibit high variance, forming what are known as volatility clusters. Figure 2.1 shows the squared S&P 500 returns used previously. Observe

that volatility is not constant and exhibits distinct clustering patterns.

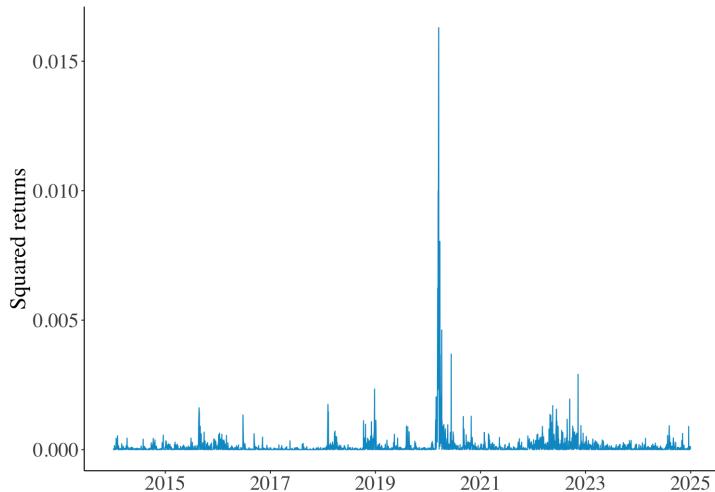


Figure 2.1: Squared S&P 500 returns, 2014-2024.

Additionally, volatility evolves continuously over time, does not diverge to infinity, and, as previously discussed, appears to respond asymmetrically to large price increases and decreases.

The literature on volatility models is extensive. Early models were developed to capture the dynamic dependency observed in variance, and subsequent models were introduced to address the limitations of their predecessors, particularly their inability to fully capture the characteristics discussed earlier. Over time, a wide range of volatility models has emerged, each catering to different modeling needs.

Before introducing specific volatility models, the squared return series, R_{t+1}^2 can be used to check for conditional heteroscedasticity. The

Ljung–Box statistic ([McLeod & Li, 1983](#)) is commonly employed for this purpose, testing the null hypothesis that the first m lags of the autocorrelation function (ACF) of the squared series are zero. Applying this test to the squared S&P 500 returns with $m = \ln(T) \approx 7$ ¹, the p-value is close to zero.

Two volatility models will be considered: the GARCH(1,1) model of [Bollerslev \(1986\)](#) and the GJR-GARCH model of [Glosten, Jagannathan, and Runkle \(1993\)](#). The GARCH(1,1) model is included because, despite its simplicity, it remains the go-to volatility model for many applied researchers—not only in finance but also in broader fields. Remarkably, GARCH(1,1) has proven difficult to convincingly outperform in out-of-sample forecasting, maintaining its status as the dominant benchmark. [Bollerslev \(2023\)](#) provides a historical overview of the GARCH model’s origins and discusses its widespread adoption, emphasizing the enduring supremacy of GARCH(1,1). The GJR-GARCH model is considered because it is the simplest extension of GARCH designed to capture the leverage effect, making it a natural choice among asymmetric volatility models.

As previously mentioned, there is a volatility model for every need. Various modifications of the GARCH model have been developed to address specific features of financial data. Some, such as EGARCH and GJR-GARCH, are designed to capture the leverage effect, while others, like IGARCH, aim to better describe long-run dynamic dependencies. Additional extensions include models that incorporate explanatory variables or integrate financial theory, such as the GARCH-M model.

¹ Setting $m = \ln(T)$ has been found to work well in simulation studies, as noted by [Christoffersen \(2012\)](#).

Many other variations exist, each tailored to different modeling objectives. Another class of volatility models not considered here is stochastic volatility models. Compared to GARCH models, they offer greater flexibility but are more challenging to estimate. [Bollerslev \(2010\)](#) provides a comprehensive guide to GARCH extensions, and [Tsay \(2010\)](#) and [Christoffersen \(2012\)](#) each dedicate a chapter in their books to volatility modeling, offering thorough reviews of the literature.

The fundamental idea in volatility modeling is that the return series R_{t+1} is either serially uncorrelated or exhibits only minor lower-order serial correlations, yet it remains dependent through its conditional variance. As noted by [Tsay \(2010\)](#), it is the time evolution of σ_{t+1}^2 that distinguishes one volatility model from another.

Recall that the return series follows [1.2](#). In the GARCH(1,1) model, the conditional variance evolves as

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2 \quad (2.1)$$

with $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$. For the GJR-GARCH(1,1) model, the conditional variance is:

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \alpha \theta I_t R_t^2 + \beta \sigma_t^2 \quad (2.2)$$

with $\omega > 0$, $\alpha \geq 0$, $\theta \geq 0$, $\beta \geq 0$, $\alpha(1+\frac{\theta}{2})+\beta < 1$, where $I_t = \mathbf{1}\{R_t < 0\}$. A positive θ captures the leverage effect.

To estimate the parameters of each model, we must assume a distribution for z_{t+1} [1.2](#). However, we can set aside this assumption for now, as the modeling of z_{t+1} will be discussed in the next section. For the time being, we will use quasi-maximum likelihood estimation (QMLE), which

refers to the application of normal maximum likelihood estimation even when the normality assumption does not necessarily hold.²

Figures 2.2 and 2.3 display the squared S&P 500 returns from Figure 2.1, now with the estimated variance from the GARCH(1,1) and GJR-GARCH(1,1) models superimposed, respectively. The parameters were estimated using the full sample (2014–2024), and the figures show the model’s performance over this same estimation period. Out-of-sample forecasting and its evaluation will be discussed in Chapter 4.

Using QMLE, the optimal parameter estimates yield the following variance dynamics for the GARCH(1,1):

$$\begin{aligned}\sigma_{t+1}^2 &= \omega + \alpha R_t^2 + \beta \sigma_t^2 \\ &= 0.0000039 + 0.171 R_t^2 + 0.794 \sigma_t^2.\end{aligned}\tag{2.3}$$

For the GJR-GARCH(1,1):

$$\begin{aligned}\sigma_{t+1}^2 &= \omega + \alpha R_t^2 + \alpha \theta I_t R_t^2 + \beta \sigma_t^2 \\ &= 0.0000040 + 0.054 R_t^2 + 0.248 I_t R_t^2 + 0.795 \sigma_t^2.\end{aligned}\tag{2.4}$$

² A key result in econometrics states that even if the conditional distribution is not normal, MLE provides consistent estimates of the mean and variance parameters as the sample size approaches infinity, provided that the mean and variance functions are correctly specified (Christoffersen, 2012). Francq and Zakoian (2010) provide a comprehensive treatment of quasi-maximum likelihood estimation (QMLE) for GARCH models, including key results such as consistency, asymptotic normality, and their proofs.

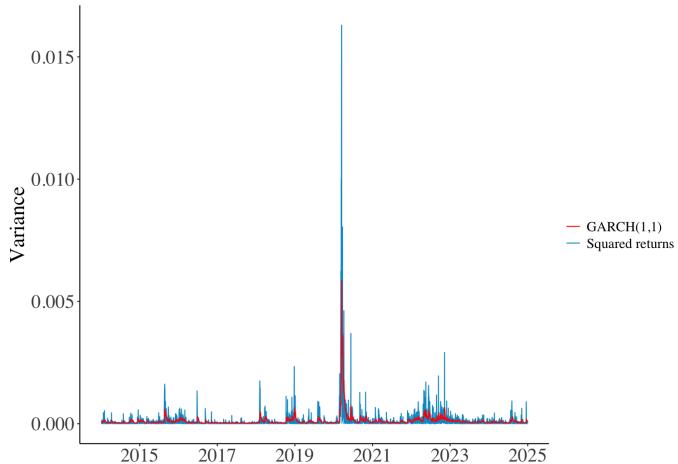


Figure 2.2: Squared S&P 500 returns with GARCH(1,1) variance parameters estimated using QMLE.

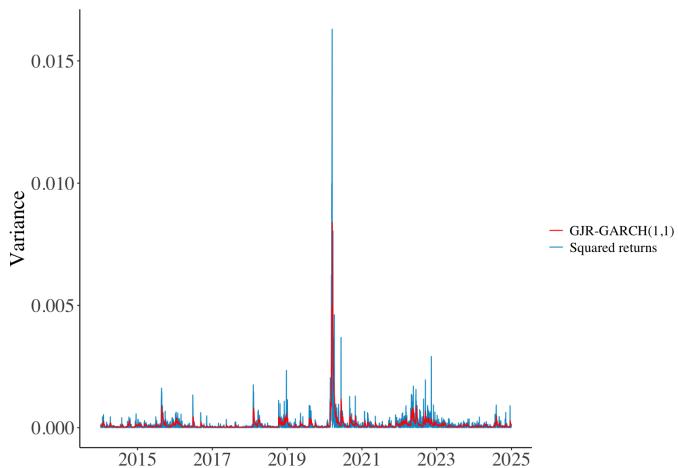


Figure 2.3: Squared S&P 500 returns with GJR-GARCH(1,1) variance parameters estimated using QMLE.

GARCH models possess several desirable theoretical properties that have been studied extensively, including the existence of a long-run variance and the ability to produce longer-horizon forecasts—features that may be relevant for risk management applications. In this chapter, we focus on daily returns and daily volatility forecasts.

Observe that $\alpha + \beta \approx 1$ in the fitted GARCH(1,1), and $\alpha + \beta + \alpha\theta \mathbb{E}[I_t] \approx 1$ in the GJR–GARCH(1,1); under symmetric innovations $\mathbb{E}[I_t] = \frac{1}{2}$, giving $\alpha + \beta + \frac{1}{2}\alpha\theta \approx 1$. This near-unit persistence is common in financial return series ([Christoffersen, 2012](#)). When the persistence parameter equals one— $\alpha + \beta = 1$ for GARCH(1,1) and $\alpha + \beta + \alpha\theta \mathbb{E}[I_t] = 1$ for GJR–GARCH(1,1)—the variance dynamics have a unit root (the IGARCH model case): the unconditional variance is not finite and multi-step variance forecasts do not mean-revert to a constant long-run level. As noted by [Tsay \(2010\)](#), $\{\sigma_t^2\}$ is a martingale in this setting, and the sources of volatility persistence warrant careful investigation.

Another important theoretical feature of GARCH models is their ability to generate return distributions with heavier tails than the normal distribution, even when normality is assumed for z_{t+1} . This means the volatility model itself contributes to capturing the excess kurtosis commonly observed in financial returns. [Bai, Russell, and Tiao \(2003\)](#) derive the kurtosis of GARCH models and show that, provided the kurtosis exists, the distribution of returns under a GARCH process can indeed exhibit heavier tails than the normal distribution.

Having analyzed return volatility, the next step is to properly model z_{t+1} , the standardized return or shock—keeping in mind the ultimate goal: to capture and replicate the empirical behavior of returns. The evaluation of the volatility models will be deferred to [Chapter 4](#), once the complete model for the return R_{t+1} has been specified.

Chapter 3

Shock Modeling

The focus now shifts to modeling the innovation z_{t+1} in 1.2. The heteroscedastic models introduced in the previous chapter help capture the excess kurtosis observed in returns. However, the distribution of the standardized return, $z_{t+1} = \frac{R_{t+1}}{\sigma_{t+1}}$, remains a challenge, as the assumption of normality is not necessarily appropriate.

Figures 3.1 and 3.2 display histograms of returns standardized using the conditional volatility from the GARCH(1,1) and GJR-GARCH(1,1) models, respectively. The empirical distribution of standardized returns continues to deviate from the standard normal distribution, exhibiting heavy tails, a degree of asymmetry, and a more pronounced peak around zero.¹

¹ A Kolmogorov–Smirnov test performed on both sets of standardized returns strongly rejects the null hypothesis of standard normality, with p-values below 0.005%. For an extensive survey of the approaches used in financial econometrics to assess and model non-normality in returns, see [Pagan \(1996\)](#).

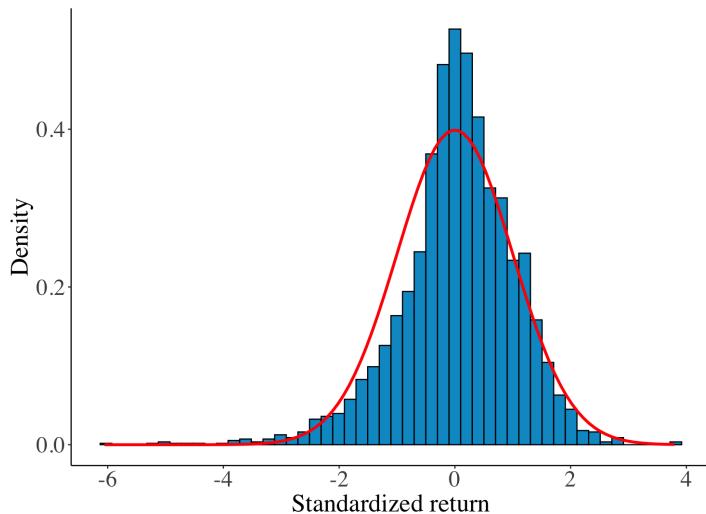


Figure 3.1: Histogram of standardized returns, GARCH(1,1).

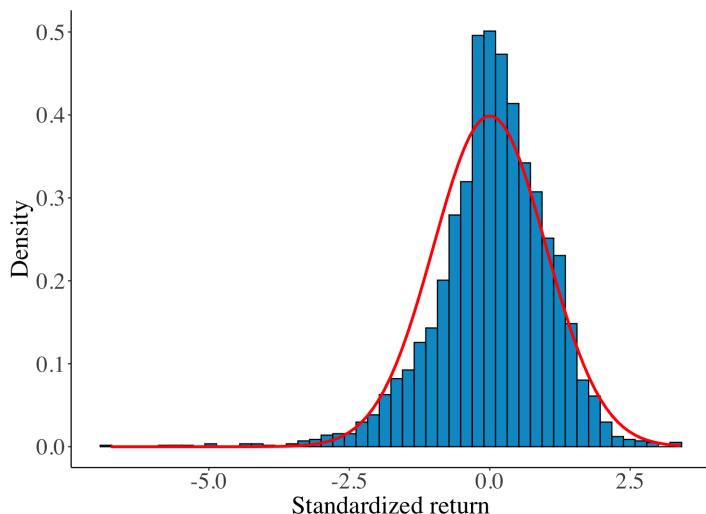


Figure 3.2: Histogram of standardized returns, GJR-GARCH(1,1).

To further illustrate the limitations of the normality assumption, consider the normal QQ plots of both sets of residuals shown in Figures 3.3 and 3.4. A particularly important weakness is the inability of the normal distribution to capture the left tail of the empirical distribution, which is considerably heavier. This has critical implications for risk management, as underestimating the probability of extreme negative returns leads to an understatement of market risk.

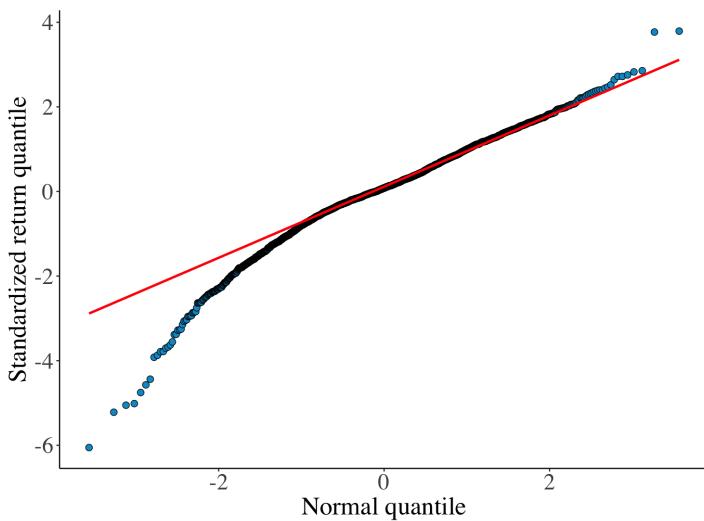


Figure 3.3: Normal QQ plot of daily standardized S&P 500 returns,
GARCH(1,1).

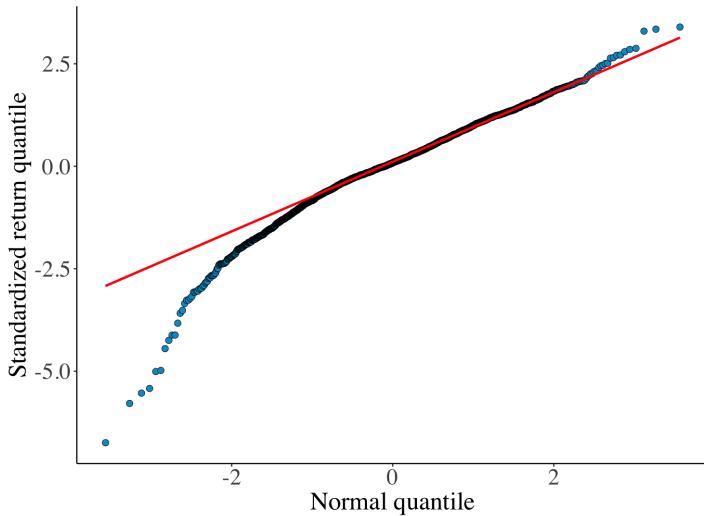


Figure 3.4: Normal QQ plot of daily standardized S&P 500 returns,
GJR-GARCH(1,1).

The objective is to identify a distribution that adequately fits the standardized returns, ensuring the model specified in Equation 1.2, with the assumption $z_{t+1} \sim \text{i.i.d. } D(0, 1)$, is appropriate. To overcome the limitations of the standard normal distribution, two alternative approaches are considered: an asymmetric extension of the standardized Student's t-distribution, and extreme value theory.

As [Cont \(2001\)](#) states, fitting various functional forms to the distribution of stock returns and stock price changes has become a popular pastime. Numerous parametric models have been proposed in the literature, starting with the normal distribution and extending to stable distributions, Student's t-distribution², and hyperbolic

²[Bollerslev \(1987\)](#) introduced an extension to the GARCH and ARCH models by

distributions, among others. Given the empirical features discussed in Chapter 1, a parametric distribution capable of capturing all these properties should have at least four parameters: a location parameter, a scale (volatility) parameter, a parameter governing tail behavior, and an asymmetry parameter allowing distinct behaviors for the left and right tails. The first of the approaches considered here—the asymmetric standardized t-distribution—meets these criteria.

The asymmetric standardized t-distribution, alternatively derived by Fernández and Steel (1998), provides a skewed extension of the Student's t-distribution, enabling the modeling of both kurtosis and skewness observed in standardized returns. The probability density function (PDF) is defined as follows:

$$f_{\text{asyt}}(z; d_1, d_2) = \begin{cases} BC \left[1 + \frac{(Bz + A)^2}{(1 - d_2)^2(d_1 - 2)} \right]^{-\frac{1+d_1}{2}}, & \text{if } z < -A/B \\ BC \left[1 + \frac{(Bz + A)^2}{(1 + d_2)^2(d_1 - 2)} \right]^{-\frac{1+d_1}{2}}, & \text{if } z \geq -A/B \end{cases} \quad (3.1)$$

where

$$A = 4d_2 C \frac{d_1 - 2}{d_1 - 1}, \quad B = \sqrt{1 + 3d_2^2 - A^2}, \quad C = \frac{\Gamma\left(\frac{d_1+1}{2}\right)}{\Gamma\left(\frac{d_1}{2}\right)\sqrt{\pi(d_1 - 2)}},$$

and where $d_1 > 2$, and $-1 < d_2 < 1$. Hansen (1994) demonstrates that this is a proper density function and applies it to the U.S. Dollar/Swiss Franc exchange rate. A random variable following this distribution has zero mean, unit variance, and skewness and kurtosis that are nonlinear functions of d_1 and d_2 (Christoffersen, 2012).

allowing for conditionally t-distributed errors.

Thus, the first complete univariate model for daily returns is defined as:

$$R_{t+1} = \sigma_{t+1} z_{t+1} \quad \text{where } z_{t+1} \sim \text{i.i.d. asyt}(d_1, d_2), \quad (3.2)$$

with conditional variance σ_{t+1}^2 given by either the GARCH(1,1):

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2, \quad (3.3)$$

or the GJR-GARCH(1,1):

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \alpha \theta I_t R_t^2 + \beta \sigma_t^2. \quad (3.4)$$

[Christoffersen \(2012\)](#) discusses various methods for estimating the model given by 3.2. In the approach followed here, the volatility parameters are first estimated using quasi-maximum likelihood estimation, and then the parameters d_1 and d_2 are estimated by maximizing the likelihood under the asymmetric Student's t-distribution.

Once the parameters of the model have been estimated, the goodness-of-fit can be evaluated visually and compared with the results obtained under the normality assumption. Figures 3.5 and 3.6 display histograms of standardized returns from the GARCH(1,1) and GJR-GARCH(1,1) models, respectively, along with their fitted asymmetric Student's t-distributions. These figures show that the asymmetric t-distribution better captures both the pronounced central peak and the skewness present in standardized returns. Additionally, the tails of the fitted distribution are heavier than those of the normal distribution, offering a closer replication of the empirical characteristics.

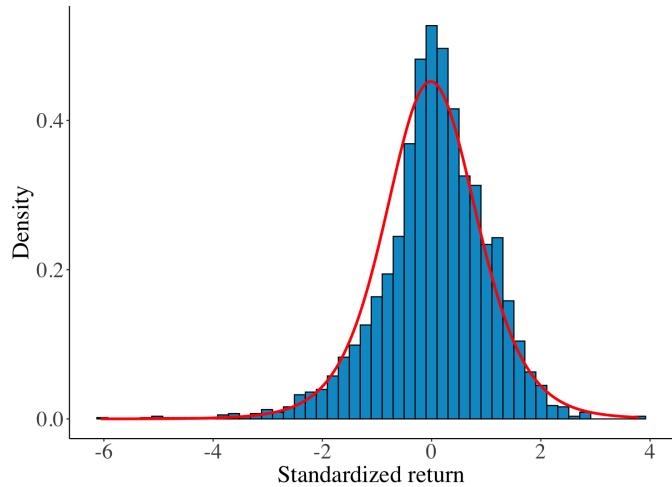


Figure 3.5: Histogram of standardized returns with the fitted asymmetric t-distribution, GARCH(1,1).

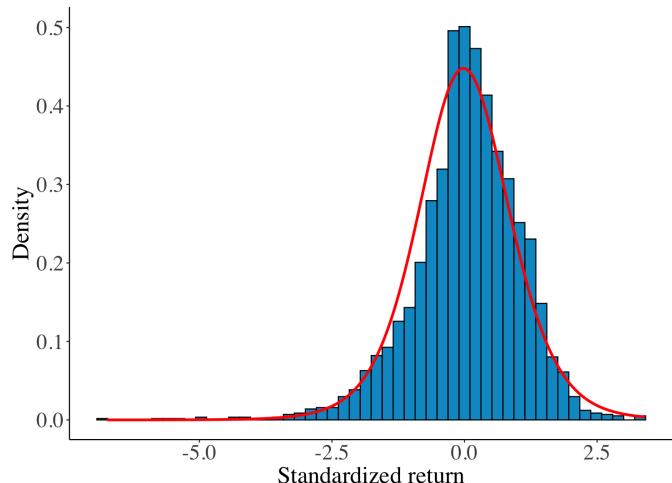


Figure 3.6: Histogram of standardized returns with the fitted asymmetric t-distribution, GJR-GARCH(1,1).

The Q–Q plots shown in Figures 3.7 and 3.8 further confirm that the asymmetric t-distribution provides a better fit to the standardized returns than the normal distribution. Nevertheless, the left tail of the empirical distribution remains inadequately modeled, exhibiting significantly heavier tails than those captured by the asymmetric t-distribution. Additionally, the comparisons between the empirical cumulative distribution functions (CDF) and their theoretical counterparts in Figures 3.9 and 3.10 reinforce the conclusion that the asymmetric t-distribution performs reasonably well in fitting the standardized returns.

Parametric models attempting to fit the full distribution are inevitably influenced by the high density of observations around the center, which may reduce their accuracy in capturing extreme values. For risk management, this is particularly concerning, as it risks underestimating the severity and likelihood of extreme negative returns. The inadequately modeled left tail demands special attention; thus, extreme value theory, which explicitly targets tail behavior, is introduced in the next approach.

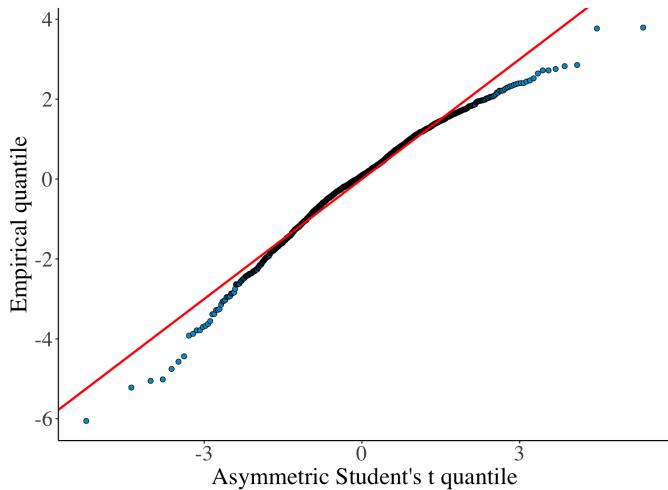


Figure 3.7: QQ plot of standardized returns against the asymmetric t-distribution, GARCH(1,1).

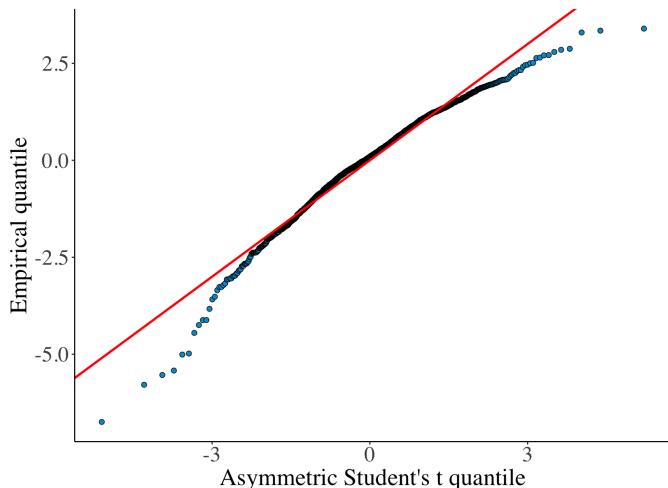


Figure 3.8: QQ plot of standardized returns against the asymmetric t-distribution, GJR-GARCH(1,1).

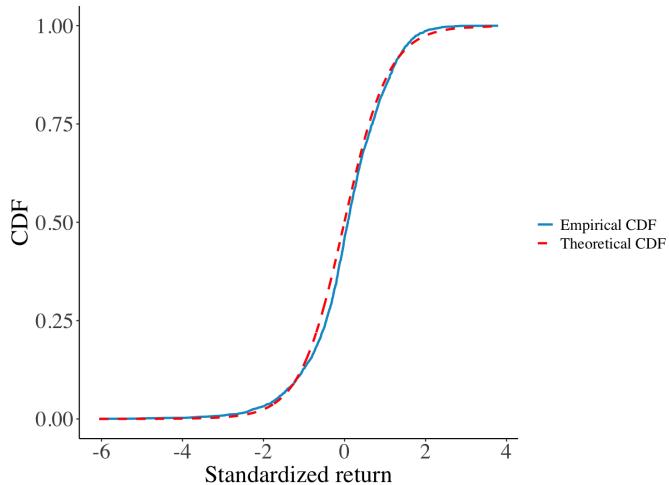


Figure 3.9: Emprical CDF and theoretical asymmetric Student's t CDF of standardized returns, GARCH(1,1).

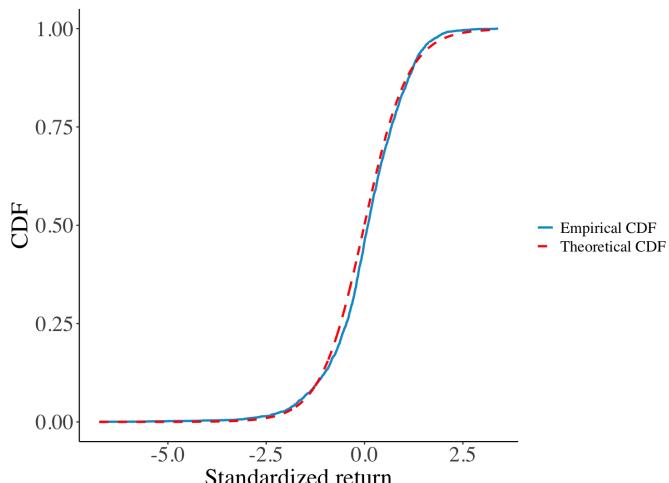


Figure 3.10: Emprical CDF and theoretical asymmetric Student's t CDF of standardized returns, GJR-GARCH(1,1).

As noted by [Christoffersen \(2012\)](#), the greatest risk to a portfolio arises from the sudden occurrence of a single large negative return. Therefore, explicit knowledge of the probabilities associated with such extreme events is central to financial risk management. Accurately modeling the left tail of the return distribution is thus critically important. Fortunately, a branch of statistics specifically addresses the modeling of these extreme events—extreme value theory (EVT).

Virtually all results in extreme value theory assume data to be independently and identically distributed (i.i.d.). Therefore, to apply this approach, the analysis is performed on standardized returns, which are assumed to meet the i.i.d. condition.

The theoretical foundation of the approach considered here is provided by the Pickands–Balkema–de Haan theorem. This theorem states that, for a wide class of distributions $X \sim F$, given a sufficiently high threshold u , the conditional distribution of excesses $(X - u)$, given $X > u$, can be approximated by a generalized Pareto distribution (GPD). For a detailed statement and an outline of the proof, see ([Cont, 2001](#)). Consequently, the GPD naturally arises as a suitable model for the conditional tail distribution of returns, as emphasized by [Tsay \(2010\)](#). Although the theorem is originally stated for the right tail, modeling the left tail simply involves applying the theorem to the negative of the returns (multiplying returns by -1).

This methodology, rooted in the Pickands–Balkema–de Haan theorem, is known in the extreme value literature as the peak-over-threshold (POT) approach. For alternative EVT-based approaches tailored specifically to financial returns, readers are referred to the chapters on extreme value theory in [Tsay \(2010\)](#) and [Christoffersen \(2012\)](#). Further applications of

EVT to financial risk management are discussed by [McNeil \(1999\)](#). The particular implementation considered here was introduced by [McNeil and Frey \(2000\)](#).

For negative standardized returns, define $y_{t+1} = -z_{t+1}$. Then, the conditional PDF for observations exceeding a threshold u is given by:

$$f(y; \xi, \beta, u | y > u) = \begin{cases} \frac{1}{\beta} \left(1 + \frac{\xi(y-u)}{\beta}\right)^{-\left(1+\frac{1}{\xi}\right)}, & \text{for } \xi \neq 0, y \geq u \text{ if } \xi > 0, \\ & u \leq y \leq u - \frac{\sigma}{\xi} \text{ if } \xi < 0; \\ \frac{1}{\beta} \exp\left(-\frac{y-u}{\beta}\right), & \text{for } \xi = 0, y \geq u. \end{cases} \quad (3.5)$$

In the above expression, ξ is the tail-index parameter, which controls the tail's heaviness. Because financial returns exhibit tails heavier than those of the normal distribution, only the case $\xi > 0$ is considered here. In this scenario, the conditional PDF simplifies to:

$$f(y; \xi, \beta, u | y > u) = \begin{cases} \frac{1}{\beta} \left(1 + \frac{\xi(y-u)}{\beta}\right)^{-\left(1+\frac{1}{\xi}\right)}, & \text{for } \xi \neq 0, y \geq u; \\ \frac{1}{\beta} \exp\left(-\frac{y-u}{\beta}\right), & \text{for } \xi = 0, y \geq u. \end{cases}$$

The second complete univariate model is therefore defined as:

$$\begin{aligned} R_{t+1} &= \sigma_{t+1} z_{t+1}, \quad \text{where } z_{t+1} \sim \text{i.i.d. } D(0, 1), \\ &\text{with } -z_{t+1} | (-z_{t+1} > u) \sim \text{GPD}(\xi, \beta, u), \end{aligned} \quad (3.6)$$

and where the conditional variance σ_{t+1}^2 follows either a GARCH(1,1) or GJR-GARCH(1,1) specification.

For parameter estimation, the approach of [McNeil and Frey \(2000\)](#) is followed. In the first step, the parameters of the volatility specification are estimated using quasi-maximum likelihood. In the second step, the parameters of the GPD are estimated via maximum likelihood using the standardized returns. [Christoffersen \(2012\)](#) discusses alternative estimation methods, including a semiparametric approach that is popular in the literature: the Hill estimator for the tail index ξ .

Before estimating the model, a threshold u must be selected. As noted by [Christoffersen \(2012\)](#), the choice of u involves a trade-off between bias and variance: a high threshold is supported by extreme value theory but may result in too few observations, increasing variance; conversely, a low threshold provides more data but may lead to overfitting and to model misspecification by fitting observations outside the tail region. Discussions on the selection of u can be found in [Tsay \(2010\)](#) and [Coles \(2001\)](#).³

Fortunately, a useful property of the generalized Pareto distribution can guide this choice. For any $u > u_0$, where u_0 is a threshold above which the GPD provides a valid approximation, the mean excess function $e(u) = \mathbb{E}[y - u \mid y > u]$ is linear in u . Thus, plotting the empirical mean excess function over a range of threshold values allows identification of the region where the function becomes approximately linear—indicating a suitable choice for u . Figures 3.11 and 3.12 suggest that choosing $u \geq 1.8$ is reasonable for both sets of standardized returns.

³[Christoffersen \(2012\)](#) also suggests selecting u by identifying the point at which the tail begins to depart from normality in the Q–Q plot of standardized residuals, and recommends using at least 50 observations in the tail fit, based on evidence from simulation studies.

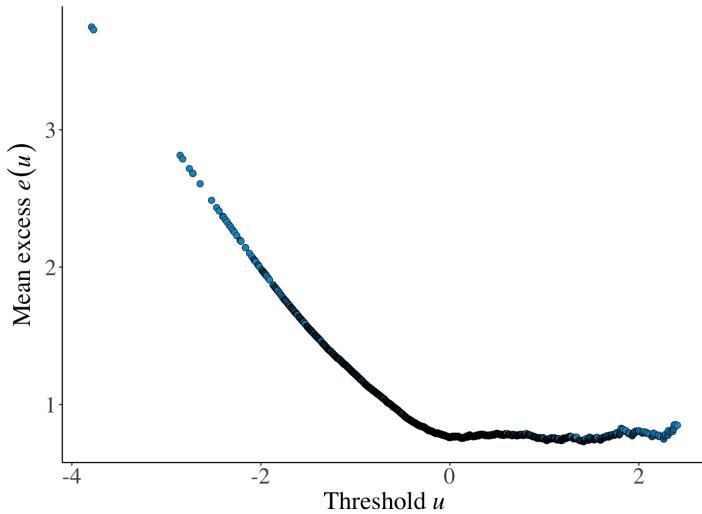


Figure 3.11: Mean excess function $e(u)$, GARCH(1,1).

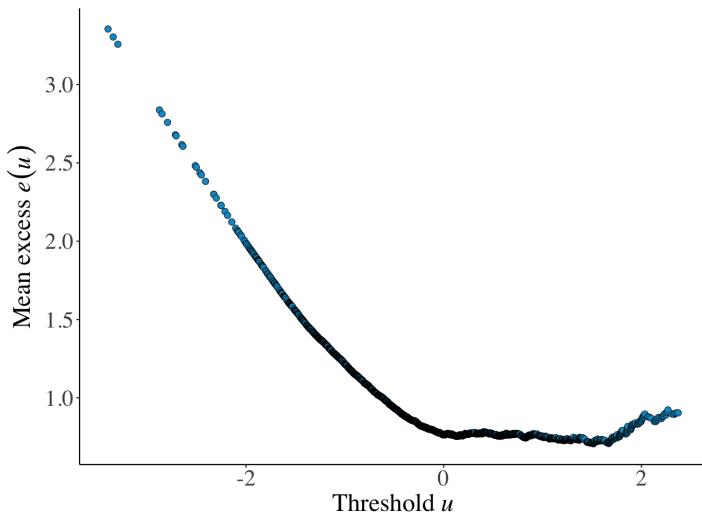


Figure 3.12: Mean excess function $e(u)$, GJR-GARCH(1,1).

Setting $u = 1.8$, the model in Equation 3.4 can be estimated, and the goodness-of-fit can be assessed visually. Figures 3.13 and 3.14 show that, for both sets of standardized returns—GARCH(1,1) and GJR-GARCH(1,1), respectively—the GPD provides a reasonable fit to the left tail.⁴ However, some nuances remain, as the model's ability to capture the tail behavior tends to diminish further into the extreme tail. To complement the visual assessment, Figures 3.15 and 3.16 display the empirical and theoretical cumulative distribution functions for both sets of standardized returns.

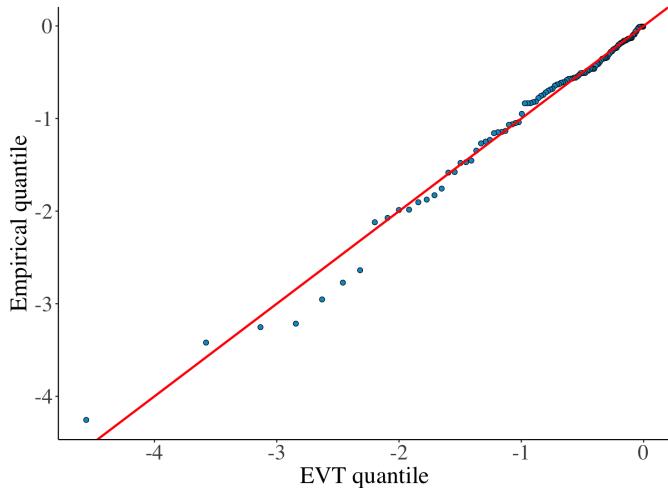


Figure 3.13: QQ plot of standardized returns against the generalized Pareto distribution, GARCH(1,1).

⁴The plots show the quantiles of the excesses over the threshold, along with the fitted GPD starting at zero, to aid in the visual assessment.

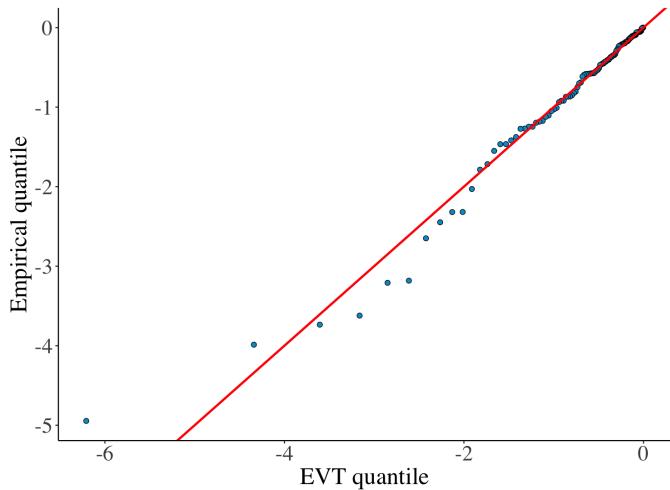


Figure 3.14: QQ plot of standardized returns against the generalized Pareto distribution, GJR-GARCH(1,1).

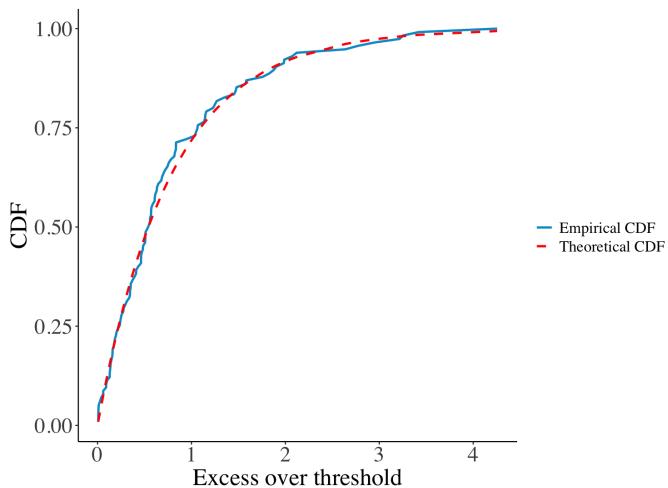


Figure 3.15: Empirical CDF and theoretical EVT CDF of standardized returns, GARCH(1,1).

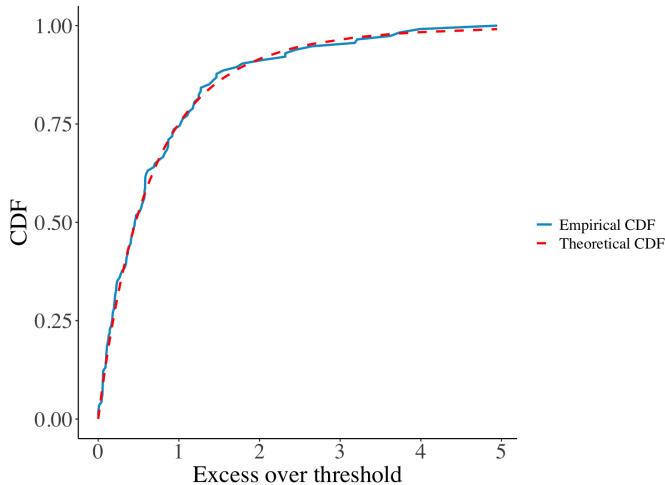


Figure 3.16: Empirical CDF and theoretical EVT CDF of standardized returns, GJR-GARCH(1,1).

This chapter concludes by summarizing the two approaches explored for modeling the standardized returns: the asymmetric Student's t-distribution and the peak-over-threshold method derived from extreme value theory. The former is a full-distribution model aimed at capturing the entire range of return behavior, while the latter focuses exclusively on the left tail to address the shortcomings of the full-distribution approach in modeling extreme losses. In essence, the first approach provides a complete density specification, whereas the second concentrates solely on tail behavior.

The variance parameters were estimated via QMLE, after which the focus shifted to the standardized returns. In total, four models for returns were considered. Table 3.1 reports the estimated parameters for each model.

Model	ω	α	β	θ	d_1	d_2	β_{GPD}	ξ
GARCH-Asyt	3.9e-6	0.171	0.794	-	7.35	0.0115	-	-
GJR-Asyt	4.0e-6	0.054	0.795	4.62	7.83	0.0129	-	-
GARCH-EVT	3.9e-6	0.171	0.794	-	-	-	0.774	0.029
GJR-EVT	4.0e-6	0.054	0.795	4.62	-	-	0.621	0.206

Table 3.1: Estimated parameters.

In the next section, the models will be used for forecasting and evaluated from both a statistical and a risk management perspective.

Chapter 4

Forecasting and Evaluation

The next step is to use the models constructed in the previous chapters for forecasting and to evaluate their performance. George E. P. Box is famously quoted as saying, “*All models are wrong, but some are useful.*” In light of Box’s observation, the task now is to evaluate whether the models constructed are among those that are useful. The evaluation will proceed from two perspectives: a statistical perspective to assess model validity, and a risk management perspective to assess their practical usefulness.

To recap and introduce the forecasting procedure, the model parameters were estimated using S&P 500 log returns from January 2014 to December 2024. The fitted models were then used to forecast the one-day-ahead variance and return density from January to April 2025. To assess their usefulness in a risk management context, the one-day-ahead Value at Risk (VaR) and Expected Shortfall (ES) are computed based on the forecasts from the fitted models.

As noted by Christoffersen (2012), before using the variance model for forecasting, it is appropriate to subject the estimated model to diagnostic checks. In Chapter 1, the dependence between squared returns was illustrated via the autocorrelation function (ACF) of squared returns. One of the primary objectives of volatility modeling is to eliminate this dependence structure by constructing a volatility measure, σ_{t+1}^2 , such that the standardized squared returns, $\frac{R_{t+1}^2}{\sigma_{t+1}^2}$, exhibit no systematic autocorrelation patterns.

Figures 4.1 and 4.2 display the ACFs of the standardized squared returns (2014-2024) under the GARCH(1,1) and GJR-GARCH(1,1) models, respectively. It can be observed that the dependence structure seen in Chapter 1 has now disappeared, indicating that the GARCH models have been effective at removing the systematic autocorrelation patterns in the squared returns.

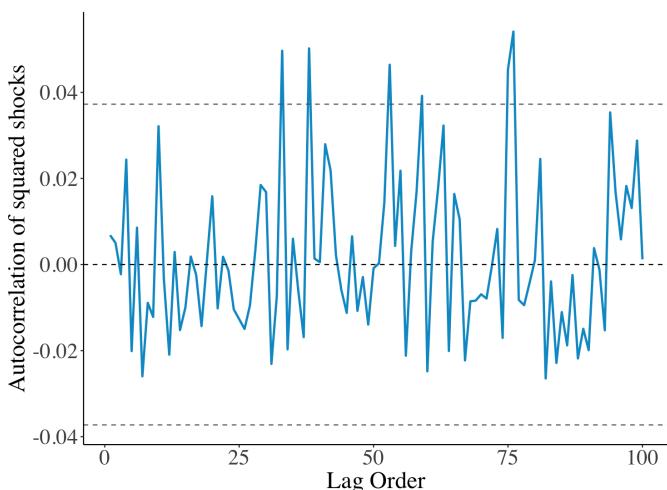


Figure 4.1: Autocorrelation of squared returns over variance (2014-2024), GARCH(1,1).

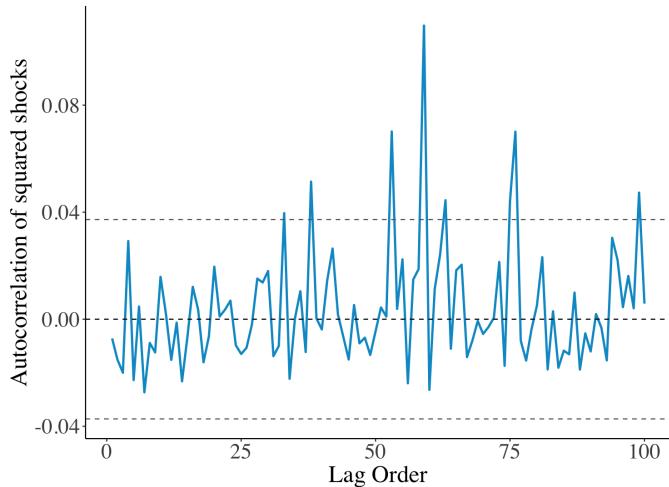


Figure 4.2: Autocorrelation of squared returns over variance (2014–2024), GJR-GARCH(1,1).

Before proceeding to forecasting, it is useful to assess which model performs better on the training set (2014–2024) under an appropriate loss function. When selecting a loss function, it is important to consider that risk managers may weigh positive and negative volatility forecast errors differently: underestimating volatility may be more costly than overestimating it by the same amount.

The QLIKE loss function addresses this asymmetry and is commonly used in the literature to evaluate volatility models, see [Elliott and Timmermann \(2016\)](#). The QLIKE loss function is defined as:

$$\text{QLIKE}_t = \frac{R_{t+1}^2}{\sigma_{t+1}^2} - \ln \left(\frac{R_{t+1}^2}{\sigma_{t+1}^2} \right) - 1. \quad (4.1)$$

A loss function is said to be robust if the ranking of any two volatility forecasts is the same when using an observed volatility proxy as when (hypothetically) using the unobserved true volatility, as defined by [Patton \(2011\)](#). Robustness is clearly a desirable property for a volatility forecast loss function, and [Patton \(2011\)](#) shows that the QLIKE loss function possesses this quality.

The mean QLIKE over the training period for both volatility models is reported in Table 4.1. The results suggest that the GJR-GARCH model outperforms the standard GARCH(1,1) model during the training period.

Model	Mean QLIKE
GARCH(1,1)	1.6011
GJR-GARCH(1,1)	1.5673

Table 4.1: Mean QLIKE over the training period

The volatility models considered here were constructed using squared returns as the volatility proxy. As noted by [Christoffersen \(2012\)](#), squared returns are an unbiased but potentially very noisy proxy for conditional variance. Alternatives include realized variance and range-based measures, which possess valuable predictive power and other desirable statistical properties. Readers interested in models based on alternative volatility proxies are referred to [Christoffersen \(2012\)](#) and the references therein. When using intraday-based proxies, it is important to account for market microstructure effects. [Tsay \(2010\)](#) and [Campbell et al. \(1997\)](#) provide overviews of the microstructure of financial markets.

A simple likelihood ratio (LR) test can also be used to assess whether the additional parameter in the GJR-GARCH model is statistically significant relative to the simpler GARCH(1,1) model. As noted by [Christoffersen \(2012\)](#), the LR test provides a straightforward way to evaluate whether added parameters are statistically meaningful.

Consider two nested models with log-likelihood values L_0 and L_1 , where model 0 is a special case of model 1. The models can be compared using the likelihood ratio statistic:

$$\text{LR} = 2(\ln(L_1) - \ln(L_0)). \quad (4.2)$$

Under the null hypothesis that the additional parameter(s) in model 1 are not significant, the LR statistic follows a chi-squared distribution with degrees of freedom equal to the number of parameters added. In this case, since GJR-GARCH(1,1) adds one parameter to GARCH(1,1), the degrees of freedom equal 1.

Using the asymmetric t-distribution as the likelihood function, the LR test yields a p-value below 1%, indicating little evidence in the data in favor of the null hypothesis. Thus, the additional parameter in the GJR-GARCH model is statistically significant.

Forecasting begins by considering the first two models in Table 3.1, which combine the volatility models with the asymmetric t-distribution. These models forecast the entire density of the next day's return, allowing for the computation of risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES). Following the definition in [Christoffersen \(2012\)](#), VaR is a risk measure that answers the question: What is the loss that

will only be exceeded $p \times 100\%$ of the time on the next trading day? In other words, VaR corresponds to a quantile of the distribution of tomorrow's return.

Under the specification $R_{t+1} = \sigma_{t+1} z_{t+1}$, with $z_{t+1} \sim \text{i.i.d. asyt}(d_1, d_2)$, the 1-day ahead p -level VaR can be computed as:

$$\text{VaR}_{t+1}^p = -\sigma_{t+1} F_{\text{asyt}}^{-1}(p; d_1, d_2), \quad (4.3)$$

where

$$F_{\text{asyt}}^{-1}(p; d_1, d_2) = \begin{cases} \frac{1}{B} \left[(1 - d_2) \sqrt{\frac{d_1 - 2}{d_1}} t_{p/(1-d_2)}^{-1}(d_1) - A \right], & \text{if } p < \frac{1 - d_2}{2} \\ \frac{1}{B} \left[(1 + d_2) \sqrt{\frac{d_1 - 2}{d_1}} t_{(p+d_2)/(1+d_2)}^{-1}(d_1) - A \right], & \text{if } p \geq \frac{1 - d_2}{2}. \end{cases}$$

The 1-day ahead Expected Shortfall (ES) is defined as

$$\text{ES}_{t+1}^p = -\mathbb{E}[R_{t+1} | R_{t+1} < \text{VaR}_{t+1}^p],$$

and under the assumption of the asymmetric t-distribution, it can be computed as:

$$\text{ES}_{t+1}^p = -\sigma_{\text{PF}, t+1} \text{ES}_{\text{asyt}}(p), \quad (4.4)$$

where

$$\begin{aligned} \text{ES}_{\text{asyt}}(p) &= \frac{C(1 - d_2)^2}{Bp} \left[1 + \frac{1}{d_1 - 2} \left(\frac{BQ + A}{1 - d_2} \right)^2 \right]^{\frac{1-d_1}{2}} \frac{d_1 - 2}{1 - d_1} \\ &\quad - \frac{AC(1 - d_2)}{Bp} \cdot \frac{\sqrt{\pi(d_1 - 2)} \Gamma\left(\frac{d_1}{2}\right)}{\Gamma\left(\frac{d_1+1}{2}\right)} t_{d_1} \left(\sqrt{\frac{d_1}{d_1 - 2}} \cdot \frac{BQ + A}{1 - d_2} \right), \end{aligned}$$

and $Q = F_{\text{asyt}}^{-1}(p; d_1, d_2)$.

According to Elliott and Timmermann (2016), a density forecast—such as those produced by combining volatility models with the asymmetric t-distribution—should satisfy certain desirable properties. One of these is *calibration*, which requires that if a density forecast assigns a specific probability to an event, then the event should occur with approximately that frequency over successive observations.

This property can be assessed using Value-at-Risk. At the $p = 2.5\%$ level, a well-calibrated model should produce VaR breaches approximately 2.5% of the time during the evaluation period. Figures 4.3 and 4.4 show the daily returns of the S&P 500 from January to April 2025, along with the one-day-ahead 2.5% VaR forecasts and the corresponding VaR breaches, assuming an asymmetric t-distribution.

With 81 observations in the test period, approximately two breaches would be expected under correct coverage. However, both models exhibit more than two breaches, indicating a failure to meet the calibration criterion. It is also worth noting the dynamic and adaptive behavior of the VaR forecast: following the market shock induced by Trump's tariff imposition, volatility increased sharply. The VaR initially fails to capture this rise—resulting in several breaches—but gradually adjusts to the new, higher volatility regime.

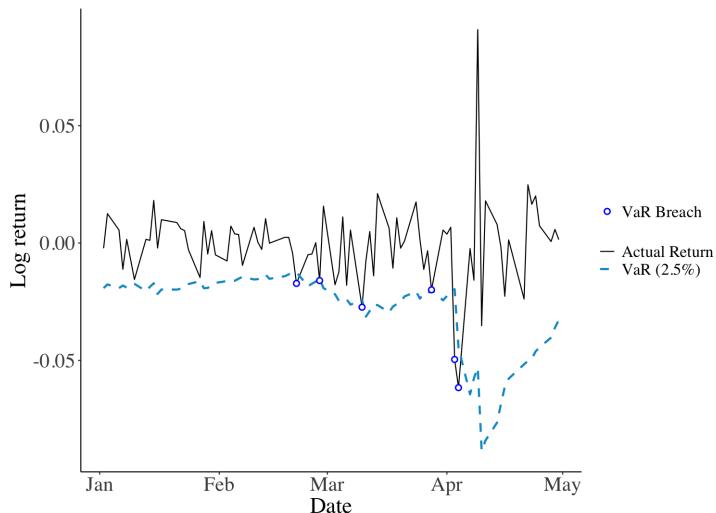


Figure 4.3: 2.5% VaR forecast and breaches in the test period, GARCH-Asyt

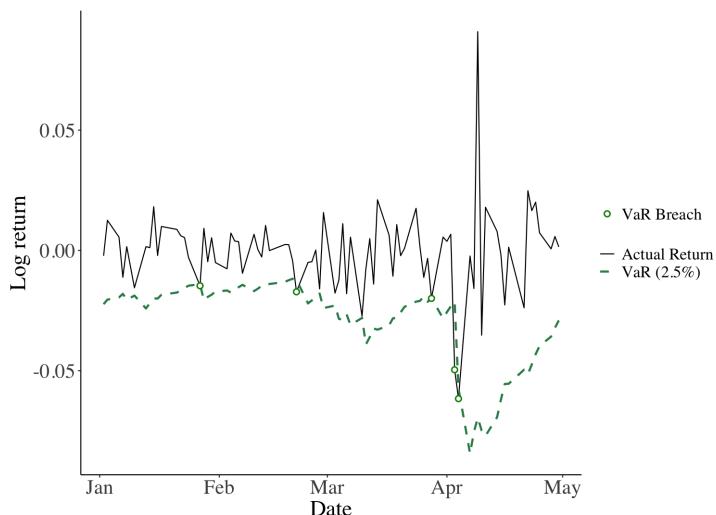


Figure 4.4: 2.5% VaR forecast and breaches in the test period, GJR-Asyt.

To improve the coverage of VaR and account for model risk, sources of uncertainty such as parameter estimation, model specification, and error distribution can be incorporated. [Mazzeu, Ruiz, and Veiga \(2018\)](#) propose several approaches to address these issues in ARMA models, which can be extended to the volatility models considered here.

As a simpler alternative, this chapter considers the use of Expected Shortfall (ES) as a way to stress the VaR forecast. Although ES is an expected value and does not correspond to a specific coverage level, it provides a more conservative risk measure by reflecting the average loss in the tail, thus functioning as a stressed version of VaR. Figures 4.5 and 4.6 display the 2.5% Expected Shortfall for the previous models.

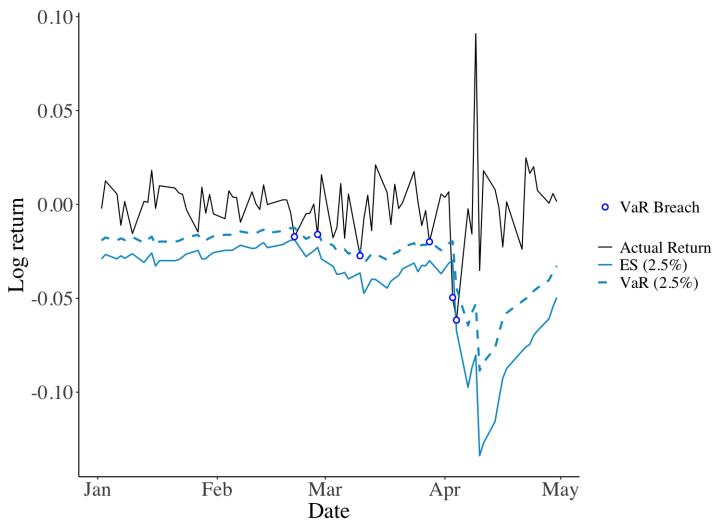


Figure 4.5: 2.5% Expected Shortfall forecast in the test period, GARCH-Asyt.

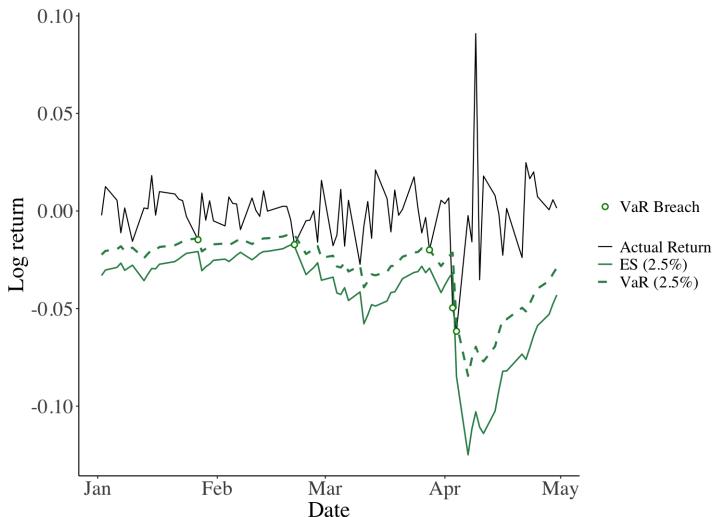


Figure 4.6: 2.5% Expected Shortfall forecast in the test period, GJR-Asyt.

Christoffersen (1998) provides an intuitive framework to evaluate the performance of VaR forecasts by testing two essential properties: coverage and independence.

Recall that a 1-day ahead VaR forecast at level p , denoted VaR_{t+1}^p , implies that the actual return will fall below this threshold only $p \times 100\%$ of the time. Given a time series of ex-ante VaR forecasts and corresponding realized returns, the hit sequence of violations can be defined as follows:

$$I_{t+1} = \begin{cases} 1, & \text{if } R_{t+1} < -\text{VaR}_{t+1}^p \\ 0, & \text{if } R_{t+1} \geq -\text{VaR}_{t+1}^p. \end{cases} \quad (4.5)$$

If the VaR model is correctly specified, violations should occur randomly

with probability p each day, implying that the hit sequence $\{I_{t+1}\}$ is independently and identically distributed (i.i.d.) as a $\text{Bernoulli}(p)$ random variable. Predictability of violations based on past information would indicate model misspecification and suggest room for improvement.

Three distinct properties can be tested, beginning with unconditional coverage. This test evaluates whether the observed frequency of breaches matches the expected frequency p . Let π represent the observed violation proportion, estimated as $\hat{\pi} = T_1/T$ (with T_1 violations observed out of T total observations). Since the Bernoulli likelihood is maximized at the sample mean, $\hat{\pi} = T_1/T$ is the maximum likelihood estimator of p . The hypotheses are stated as:

$$H_0 : I_{t+1} \sim \text{Bernoulli}(p) \quad \text{vs.} \quad H_1 : I_{t+1} \sim \text{Bernoulli}(\pi).$$

The likelihood ratio test statistic for unconditional coverage is given by $LR_{uc} = -2 \ln \left(\frac{L(p)}{L(\hat{\pi})} \right)$ which, under the null hypothesis, asymptotically follows a χ^2 distribution with 1 degree of freedom.

Second, the independence of VaR violations can be assessed. Independence is critical; even if a VaR model achieves correct unconditional coverage, clustered violations would be problematic from a risk management perspective. Ideally, violations should occur randomly and independently, rather than grouping together in time. Therefore, testing for independence involves checking if the violations exhibit temporal clustering.

To conduct this test, assume the hit sequence $\{I_t\}$ follows a first-order

Markov process characterized by the transition probability matrix:

$$\Pi_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}.$$

where, for example, the probability of a VaR violation tomorrow ($I_{t+1} = 1$) given no violation today ($I_t = 0$) is denoted by π_{01} . With a sample of T observations, the likelihood function under this first-order Markov assumption is:

$$L(\Pi_1) = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}}.$$

where T_{ij} , $i, j \in \{0, 1\}$, is the number of times is the count of times a state j occurs immediately after state i . The parameters π_{01} and π_{11} are estimated via maximum likelihood.

The null hypothesis of independence assumes that the probability of a violation does not depend on previous violations, thus simplifying the transition matrix to:

$$\Pi_0 = \begin{bmatrix} 1 - \pi & \pi \\ 1 - \pi & \pi \end{bmatrix}.$$

The likelihood ratio test for independence and its distribution under the null hypothesis is:

$$LR_{\text{ind}} = -2 \ln \left[\frac{L(\hat{\Pi})}{L(\hat{\Pi}_1)} \right] \stackrel{\text{d}}{\sim} \chi^2_1,$$

where $L(\hat{\Pi}_0)$ and $L(\hat{\Pi}_1)$ represent the maximum likelihood under the null hypothesis of independence and under the alternative hypothesis allowing dependence, respectively.

Lastly, it is critical to jointly test conditional coverage, evaluating simultaneously both independence and correct unconditional coverage.

This is done via the conditional coverage likelihood ratio test which, under the null hypothesis, asymptotically follows a χ^2 distribution with 2 degrees of freedom:

$$LR_{cc} = -2 \ln \left[\frac{L(p)}{L(\hat{\Pi}_1)} \right] \dot{\sim} \chi^2_2.$$

which decomposes neatly into two previously described components:

$$LR_{cc} = LR_{uc} + LR_{ind}.$$

In practice, backtesting frequently suffers from limited data, resulting in a relatively small number of violations T_1 , which are crucial for the tests. Thus, relying on asymptotic χ^2 distributions may not always be appropriate. Monte Carlo simulated p-values provide a robust alternative in this scenario. [Christoffersen and Pelletier \(2004\)](#) discuss the implementation of such Monte Carlo-based p-values, a method originally proposed by [Dufour \(2006\)](#).

To calculate these simulated p-values, generate 999 random samples of i.i.d. Bernoulli(p) variables, each of length T , matching the actual sample size. From these artificial samples, compute 999 simulated test statistics, denoted as $\{\widetilde{LR}_{uc}(i)\}_{i=1}^{999}$. Finally, the simulated p-value is calculated as the proportion of simulated LR_{uc} statistics exceeding the observed test statistic, providing a reliable measure of the test's significance.

Table 4.2 reports the theoretical and simulated p-values for the tests proposed by [Christoffersen \(1998\)](#), applied to the GARCH-Asyt and GJR-Asyt models.

Test	p-value	GARCH-Asyt	GJR-Asyt
LR_{uc}	Theoretical	0.0215	0.0735
	Simulated	0.003	0.144
LR_{ind}	Theoretical	0.4332	0.2793
	Simulated	0.095	0.047
LR_{cc}	Theoretical	0.0523	0.1123
	Simulated	0.015	0.028

Table 4.2: Theoretical and simulated p-values for VaR backtesting statistics.

These models exhibit weak performance: they fail to achieve acceptable coverage, and the violations do not seem independent. As shown in Figures 4.3 and 4.4, two breaches occur in quick succession at the onset of President Trump’s tariff imposition, indicating slow adaptation to a sharp increase in volatility and resulting in clustered VaR violations.

Attention now turns to evaluating the quality of the density forecasts. Focusing on the full forecasted return distribution offers the advantage of potentially increasing the power to reject misspecified risk models.

The GARCH-Asyt and GJR-Asyt models generate a one-step-ahead cumulative distribution forecast for the next day’s return, denoted $F_t(\cdot)$, at the end of each trading day. After observing the actual return R_{t+1} , the cumulative probability assigned to this outcome by the forecasted distribution can be evaluated. This probability, known as the Probability Integral Transform (PIT), was introduced by Diebold, Gunther, and Tay

(1998) and is defined as:

$$u_{t+1} = F_t(R_{t+1}). \quad (4.6)$$

Diebold et al. (1998) show that if the risk model used to forecast the return distribution is correctly specified, then the sequence $\{u_{t+1}\}$ should be independently and identically distributed as Uniform(0,1). Intuitively, this means that if the forecasted distribution is correct, the PIT values should not be predictable or systematically biased.

To assess the validity of the risk models considered here, the hypothesis of interest is:

$$H_0 : u_{t+1} \sim \text{i.i.d. Uniform}(0, 1).$$

Diebold et al. (1998) suggest a visual diagnostic using a histogram of the PIT values $\{u_{t+1}\}$ to assess whether they appear approximately uniform. If systematic deviations from flatness are observed, this may indicate that the forecast distribution is misspecified. Figures 4.7 and 4.8 display the PIT histograms for the GARCH-Asyt and GJR-Asyt models over the test period, respectively. There are no evident deviations from uniformity.

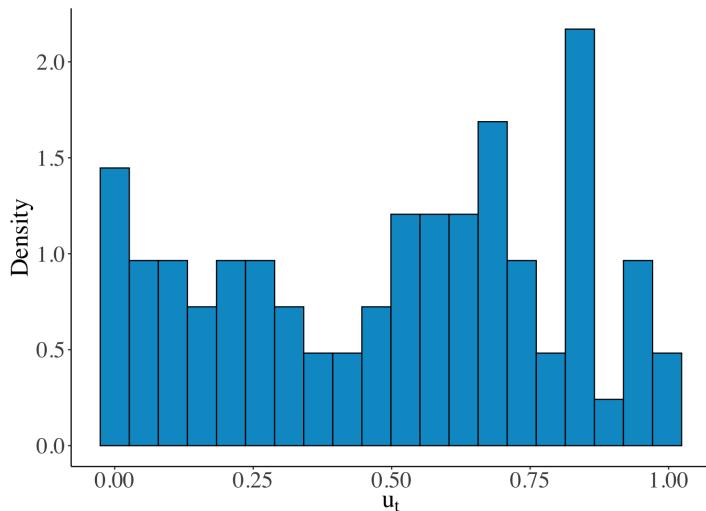


Figure 4.7: Histogram of PIT values, GARCH- Asyt.

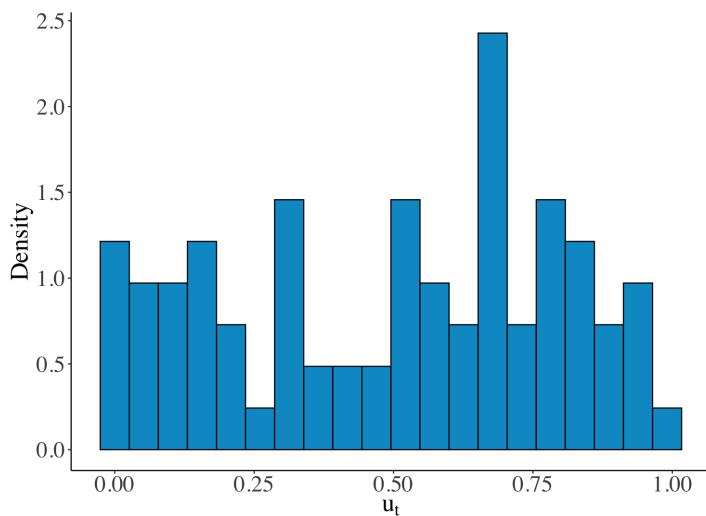


Figure 4.8: Histogram of PIT values, GJR- Asyt.

While the histogram provides an intuitive diagnostic, it is not a formal statistical test. As noted by [Christoffersen \(2012\)](#), testing the i.i.d. uniformity hypothesis is challenging due to the bounded support of the uniform distribution. To address this, [Berkowitz \(2001\)](#) proposes transforming the PIT values to a standard normal scale and applying standard statistical tests to the transformed series:

$$\hat{z}_{t+1} = \Phi^{-1}(u_{t+1}), \quad (4.7)$$

where $\Phi^{-1}(\cdot)$ denotes the inverse of the standard normal cumulative distribution function. The hypothesis of interest becomes:

$$H_0 : \hat{z}_{t+1} \sim \text{i.i.d. } N(0, 1).$$

To test this hypothesis, a Kolmogorov–Smirnov (KS) test can be applied to the GARCH-Asyt and GJR-Asyt models to assess the validity of the density forecasts. Figures 4.9 and 4.8 display the empirical cumulative distribution functions of the transformed series $\{\hat{z}_{t+1}\}$ plotted against the standard normal CDF, with the KS test statistics indicated.

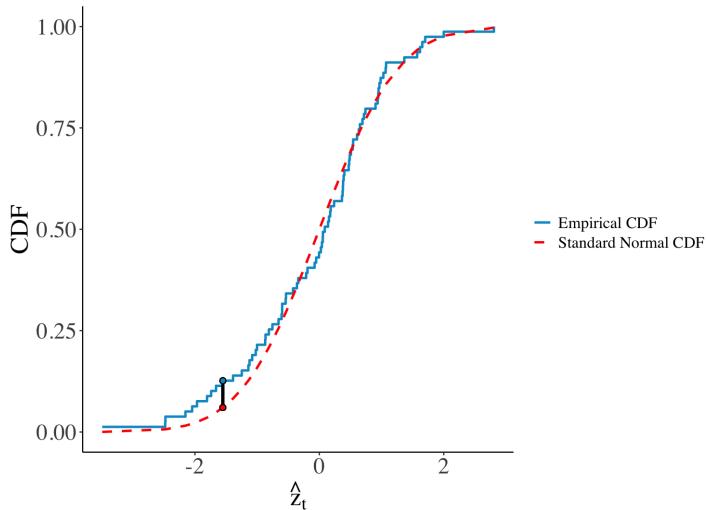


Figure 4.9: Kolmogorov-Smirnov test on \hat{z}_{t+1} , GARCH-Asyt.

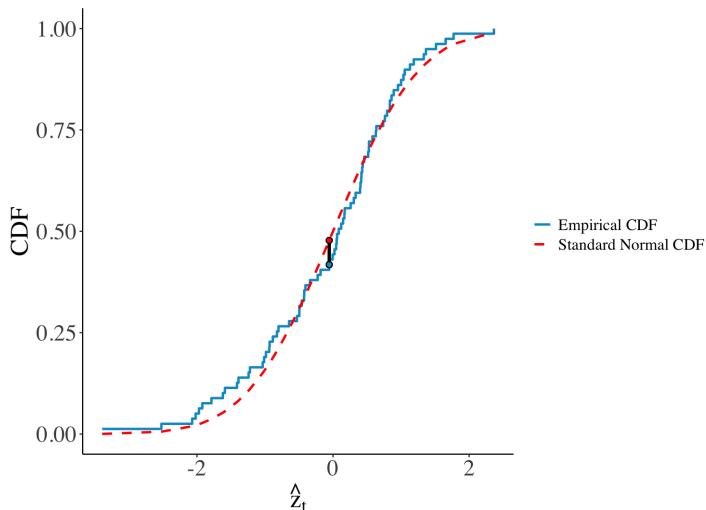


Figure 4.10: Kolmogorov-Smirnov test on \hat{z}_{t+1} , GJR- Asyt.

The resulting p -values for the GARCH-Asyt and GJR-Asyt models are 0.8512 and 0.7931, respectively. These values suggest that the hypothesis of normality cannot be rejected, supporting the validity of the models' density forecasts. In addition, [Berkowitz \(2001\)](#) proposes a likelihood ratio (LR) test to jointly assess normality and independence. Specifically, the hypothesis

$$H_0 : \hat{z}_{t+1} \sim \text{i.i.d. } N(0, 1) \quad vs. \quad H_1 : \hat{z}_{t+1} \sim N(\hat{\mu}, \hat{\sigma}^2)$$

with autocorrelation ρ , can be tested via an LR statistic. The resulting p -values for the GARCH-Asyt and GJR-Asyt models are 0.3089 and 0.6577, respectively, indicating that the null hypothesis cannot be rejected.

To formally compare the density forecasts during the test period, the Diebold–Mariano (DM) test introduced by [Diebold and Mariano \(1995\)](#) is employed. This test compares the accuracy of two competing forecast models under a specified loss function, which in this case is the negative log-likelihood function of the actual returns given the forecasted densities. Using the log-likelihood as a scoring rule follows the suggestion of [Elliott and Timmermann \(2016\)](#) for evaluating density forecasts.

The DM test statistic is defined as:

$$\text{DM} = \frac{\bar{d}}{\hat{\sigma}_d / \sqrt{T}},$$

where \bar{d} is the average difference in log-likelihood scores, $d_t = \ell^{(\text{GARCH})}(R_{t+1}|F_t) - \ell^{(\text{GJR})}(R_{t+1}|F_t)$,

with $\ell^{(m)}(\cdot)$ and $\ell^{(n)}(\cdot)$ representing the log-likelihoods of the two competing models. Here, $\hat{\sigma}_d$ denotes the standard deviation of the series

$\{d_t\}$, and T is the sample size. Under the null hypothesis of equal forecast accuracy, the DM statistic follows a standard normal distribution.

After performing the test, the resulting p -value is 0.015, with a positive test statistic, suggesting that the GJR-Asyt density forecast provides significantly better forecasting performance compared to the GARCH-Asyt model.

[Corradi and Swanson \(2006\)](#) provide an extensive survey of methods to evaluate and compare density forecasts, both in-sample and out-of-sample. Readers interested in further methods for evaluating density forecasts are referred to this comprehensive survey.

Attention now turns toward the GARCH-EVT and GJR-EVT models. These models provide conditional tail distributions rather than full-density forecasts. Informally, one might describe them as tail forecasts. Although several frameworks exist to evaluate tail forecasts, such as those proposed by [Berkowitz \(2001\)](#) and [Christoffersen \(2012\)](#), these approaches cannot be directly adapted to the conditional GARCH-EVT and GJR-EVT models considered here. Thus, these two models will only be assessed from a risk management perspective; a rigorous statistical validation method remains to be formally established. One possible validation strategy could involve adapting the PIT approach specifically for observations exceeding the threshold u , but this idea has not yet been fully explored in the reviewed literature.

According to the Pickands–Balkema–de Haan theorem, for sufficiently large threshold u ,

$$\frac{F(y) - F(u)}{1 - F(u)} \approx G_{u,\xi,\beta}(y) \quad \text{for } y > u,$$

where $G_{u,\xi,\beta}(y)$ denotes the cumulative distribution function (CDF) of the generalized Pareto distribution (GPD). Approximating the CDF $F(u)$ of the standardized returns using the empirical distribution, we have:

$$\hat{F}(\eta) = \frac{T - N_u}{T},$$

where N_u is the number of threshold exceedances and T is the training sample size. Consequently, the tail distribution can be approximated by:

$$F(y) \approx F(u) + G(y)[1 - F(u)] \approx 1 - \frac{N_u}{T} \left[1 + \frac{\xi(y - u)}{\beta} \right]^{-1/\xi}.$$

This approximation provides a direct method to estimate quantiles used for VaR calculations. Specifically, for a small tail probability p , letting $q = 1 - p$, solving the previous equation for y yields the estimated VaR of the standardized returns:

$$\text{VaR}_{z,t+1}^p = u - \frac{\beta}{\xi} \left\{ 1 - \left[\frac{T}{N_u} (1 - q) \right]^{-\xi} \right\}. \quad (4.8)$$

Consequently, the VaR of actual returns is given by:

$$\text{VaR}_{t+1}^p = -\sigma_{t+1} \text{VaR}_{z,t+1}^p. \quad (4.9)$$

Following similar logic, the Expected Shortfall (ES) for the GARCH-EVT and GJR-EVT models is defined as:

$$\text{ES}_{t+1}^p = -\sigma_{t+1} \text{ES}_{z,t+1}^p, \quad (4.10)$$

where $\text{ES}_{z,t+1}^p = \frac{\text{VaR}_{z,t+1}^p}{1-\xi} + \frac{\beta-\xi u}{1-\xi}$.

Figures 4.11 and 4.12 display the forecasted 2.5% VaR, ES, and the observed VaR breaches during the test period. Notice that these models, which incorporate extreme value theory, yield a more conservative (larger magnitude) VaR and consequently exhibit fewer breaches compared to the asymmetric t-distribution models.

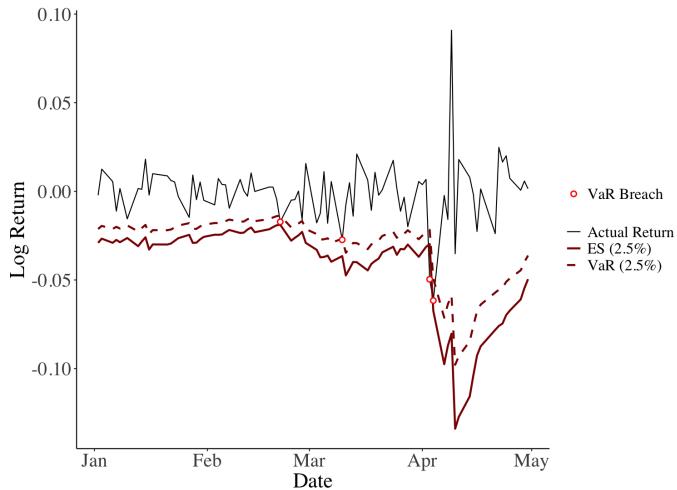


Figure 4.11: 2.5% VaR ES, and VaR breaches in the test period, GARCH-EVT.

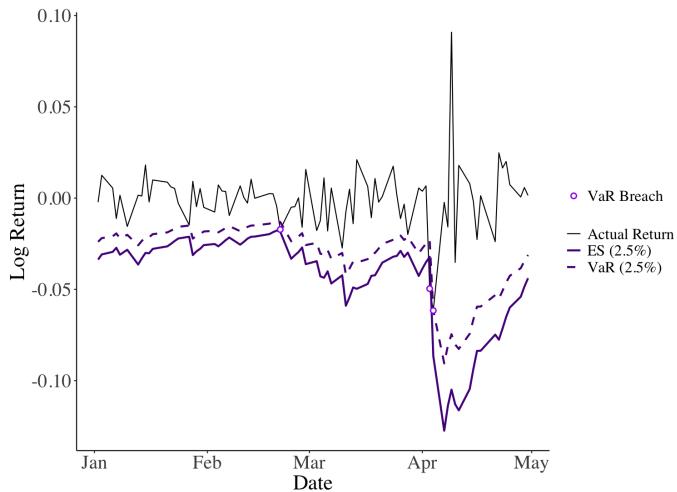


Figure 4.12: 2.5% VaR ES, and VaR breaches in the test period, GJR-EVT.

The results of the VaR tests using the methodology of [Christoffersen \(1998\)](#) are summarized in Table 4.3.

Test	p-value	GARCH-EVT	GJR-EVT
LR_{uc}	Theoretical	0.2138	0.5168
	Simulated	0.188	0.554
LR_{ind}	Theoretical	0.1582	0.0729
	Simulated	0.03	0.013
LR_{cc}	Theoretical	0.1706	0.1622
	Simulated	0.217	0.2

Table 4.3: Theoretical and simulated p -values for VaR testing statistics.

Both models offer acceptable coverage, but show weaknesses regarding the independence of VaR breaches; primarily due to the two consecutive violations at the onset of the tariff imposed by the United States. This suggests that the models respond slowly to sudden volatility increases. The GJR-EVT model provides acceptable coverage but still fails the independence test. Its comparatively better performance arises from the GJR volatility specification, which adapts more quickly to the volatility increases in the test sample (though not rapidly enough to avoid consecutive breaches), along with the more conservative tail estimation offered by the EVT approach.

Finally, as an exploratory stress-testing exercise, the four models considered (GARCH-Asyt, GJR-Asyt, GARCH-EVT, and GJR-EVT) were assessed for their performance during the 2008 financial crisis, evaluating their effectiveness from a risk management perspective. The results are presented in Figure 4.13. Notice that the EVT-based models exhibit fewer VaR breaches, with the GARCH-EVT model performing best in terms of minimizing VaR violations. Additional stress-testing methodologies and their implementation are discussed by [Christoffersen \(2012\)](#).

Up to now, only univariate models have been considered. The next chapter adopts a copula-based approach to model dependence among assets and produce multivariate forecasts. The reason behind this approach is that copulas integrate naturally with the univariate models developed so far.

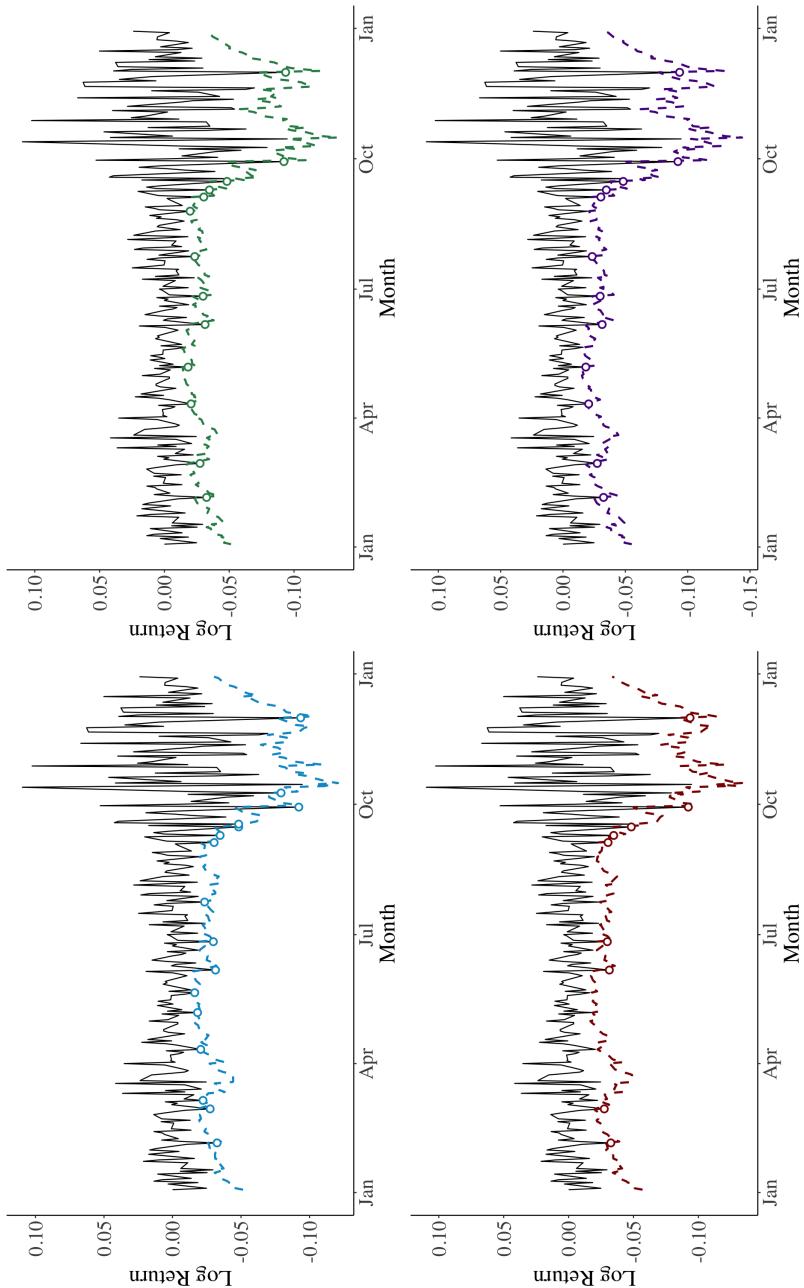


Figure 4.13: Stress test of the models.

Chapter 5

Copula Multivariate Model

The previous chapters covered various approaches to modeling univariate asset returns. These univariate methods can also be applied directly to portfolio returns. However, while modeling aggregate portfolio returns is suitable for passive risk measurement, it offers limited insight for active risk management. To conduct sensitivity analyses and evaluate diversification benefits, it is necessary to model the dependence structure between individual asset returns. This chapter discusses the copula approach for constructing multivariate risk models.

Alternative approaches to modeling a set of assets (multivariate risk) include those based on the Dynamic Conditional Correlation (DCC) model, which captures the linear dependence between standardized asset returns. A key strength of the DCC approach is its ability to reflect the dynamic correlation patterns introduced in [Chapter 1](#) and illustrated in [Figure 5.1](#). Additionally, the DCC model can be combined with multivariate distributions designed to capture asymmetry and heavy

tails—features previously discussed in a univariate context but now extended to multiple dimensions. Readers interested in exploring this approach in detail are encouraged to consult [Christoffersen \(2012\)](#), which provides a comprehensive overview of the DCC model and compatible multivariate densities.

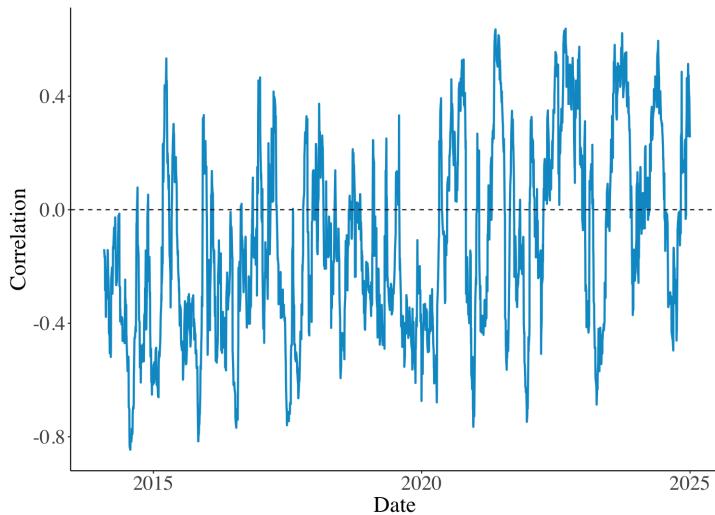


Figure 5.1: 25-Day rolling correlation of standardized returns (training period).

The DCC model approach has two key limitations. First, it focuses exclusively on capturing linear dependencies, thus overlooking more complex nonlinear relationships—particularly in the tails of asset distributions, which are critical for effective risk management. Second, it constrains the univariate models by tying their parameters directly to those estimated within the multivariate framework, potentially limiting their ability to accurately represent individual asset behavior.

The copula modeling approach addresses these issues. It can capture both linear and nonlinear dependencies effectively and allows for separate, flexible univariate models for each asset. These individual univariate models serve as inputs to the copula—analogous to ingredients being blended—resulting in a multivariate model that preserves each asset’s distinctive univariate properties.

Although copula models can be specified dynamically, this chapter employs a static copula for the standardized returns, following [Jondeau and Rockinger \(2006\)](#). In this static specification, the copula parameters are time-invariant (e.g., in a t-copula the correlation parameter ρ^* and the degrees of freedom d do not vary with t); any dynamics arise solely from the marginal models. Consequently, the specification does not capture time variation in cross-asset dependence (e.g., dynamic correlations). Dynamic copula models have been extensively developed by [Patton \(2006\)](#), [Patton and Oh \(2011\)](#), and [Christoffersen and Langlois \(2011\)](#). In particular, [Christoffersen and Langlois \(2011\)](#) embed Dynamic Conditional Correlation dynamics within the copula, yielding time-varying dependence.

From a risk management perspective, the tail dependence between assets warrants particular attention. A portfolio containing assets with strongly correlated left-tail returns could suffer significant losses during sharp market downturns. To visualize such tail dependencies, threshold correlations play a key role. The threshold correlation between two standardized returns $z_{1,t}$ and $z_{2,t}$ at level p is defined as

$$\rho(r_{1,t}, r_{2,t}; p) = \begin{cases} \text{Corr}(z_{1,t}, z_{2,t} \mid z_{1,t} \leq z_1(p) \text{ and } z_{2,t} \leq z_2(p)), & \text{if } p \leq 0.5 \\ \text{Corr}(z_{1,t}, r_{2,t} \mid z_{1,t} > z_1(p) \text{ and } z_{2,t} > z_2(p)), & \text{if } p > 0.5 \end{cases} \quad (5.1)$$

where $z_i(p)$ represents the p -th percentile of the standardized returns of asset i .

In other words, the threshold correlation measures the correlation between two assets conditioned on both returns being simultaneously below their respective p -th percentiles (for $p < 0.5$) or above these percentiles (for $p > 0.5$). Calculating threshold correlations across a range of p -values and plotting these correlations yields the threshold correlation plot. As noted by [Christoffersen \(2012\)](#), threshold correlations reveal how asset returns behave jointly during extreme events—large negative or positive returns—thus providing insights into the shape of the tails of their joint distribution.

Figure 5.2 displays the threshold correlation plot between the S&P 500 returns and the 10-year Treasury bond returns standardized by their respective GARCH(1,1) volatilities. For values of p close to 0 or 1, the number of observations becomes limited, which prevents reliable correlation estimation; thus, only correlations computed with at least 25 observations are presented. Although the extreme correlations exhibit considerable variability and should be interpreted cautiously, an important pattern emerges clearly: correlations increase as returns move deeper into the tails. This indicates a heavy-tailed bivariate distribution between stock and bond returns. Additionally, note that there is asymmetry in the tail dependence, the right tail exhibits a higher correlation.

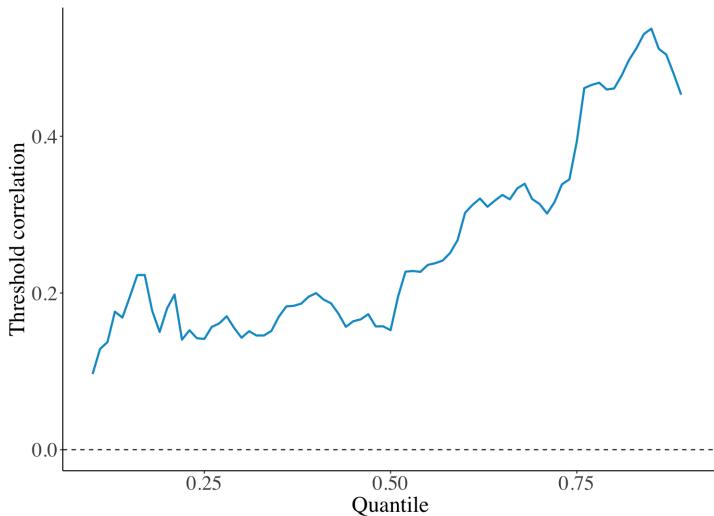


Figure 5.2: Threshold correlation of GARCH(1,1) standardized returns (training period).

Such tail dependence exemplifies the nonlinear relationships that the copula modeling approach aims to capture and replicate. The results in Figure 5.2 underscore the importance of modeling nonlinear left-tail dependencies between stocks and bonds. An influential application of threshold correlations is provided by [Okimoto \(2008\)](#), who investigates asymmetric dependence structures among international equity markets.

Following the exposition by [Christoffersen \(2012\)](#), the theoretical foundation for copula models is provided by Sklar's theorem. Consider n assets and their respective standardized returns z_i , each with potentially distinct marginal probability density functions (PDFs) $f_i(z_i)$ and cumulative distribution functions (CDFs) $u_i = F_i(z_i)$ for $i = 1, 2, \dots, n$. Sklar's theorem states that for a very general class of multivariate cumulative distribution functions $F(z_1, \dots, z_n)$ with

marginals $F_1(z_1), \dots, F_n(z_n)$, there exists a unique copula function $G(\cdot)$ that links these marginals to form the joint distribution:

$$F(z_1, \dots, z_n) = G(F_1(z_1), \dots, F_n(z_n)) = G(u_1, \dots, u_n).$$

The function $G(u_1, \dots, u_n)$ is referred to as the copula CDF. In line with Sklar's theorem, the corresponding multivariate probability density function (PDF) can be expressed as:

$$\begin{aligned} f(z_1, \dots, z_n) &= \frac{\partial^n G(F_1(z_1), \dots, F_n(z_n))}{\partial z_1 \cdots \partial z_n} \\ &= \frac{\partial^n G(u_1, \dots, u_n)}{\partial u_1 \cdots \partial u_n} \times \prod_{i=1}^n f_i(z_i) \\ &= g(u_1, \dots, u_n) \times \prod_{i=1}^n f_i(z_i), \end{aligned}$$

where the copula PDF $g(u_1, \dots, u_n)$ is defined by:

$$g(u_1, \dots, u_n) \equiv \frac{\partial^n G(u_1, \dots, u_n)}{\partial u_1 \cdots \partial u_n}.$$

Taking logarithms simplifies the representation as follows:

$$\ln f(z_1, \dots, z_n) = \ln g(u_1, \dots, u_n) + \sum_{i=1}^n \ln f_i(z_i).$$

This decomposition demonstrates that constructing a complex multivariate density can be performed in two steps. First, individual marginal distributions $f_1(z_1), \dots, f_n(z_n)$ are estimated independently using methods outlined in previous chapters. Second, the copula density function $g(u_1, \dots, u_n)$ is chosen and estimated using the marginal probability values u_i as input data.

An important practical advantage of this method is that it simplifies parameter estimation: the parameters of the marginal distributions

are estimated independently first, reducing the copula estimation to a separate, simpler second step. This property makes high-dimensional modeling feasible and efficient.

While Sklar's theorem is quite general, holding true for a broad class of multivariate distributions, it does not specify a functional form for $G(u_1, \dots, u_n)$. Therefore, an appropriate copula model must be selected based on the application and underlying data.

The t-copula, widely used in the literature (see, for instance, [Christoffersen and Langlois \(2011\)](#)), can be extended naturally to n assets and, as will be demonstrated subsequently, is capable of reproducing reasonably the threshold correlations shown in Figure 5.2. For these reasons, the t-copula is selected for our analysis. However, it is important to highlight that [Okimoto \(2008\)](#) finds evidence of significant asymmetric dependence structures in financial data, which symmetric copulas—such as Gaussian and t-copulas—may fail to fully capture.

The cumulative distribution function of the bivariate t-copula is defined as:

$$G(u_1, u_2; \rho^*, d) = t_{(d, \rho^*)}(t^{-1}(u_1; d), t^{-1}(u_2; d)) \quad (5.2)$$

where $t_{(d, \rho^*)}(\cdot)$ denotes the (non-standardized) symmetric multivariate t-distribution, $t^{-1}(u; d)$ is the inverse CDF of the (non-standardized) symmetric univariate t-distribution, and ρ^* is the correlation parameter between $t^{-1}(u_1; d)$ and $t^{-1}(u_2; d)$, commonly known as the copula correlation.

The corresponding probability density function or the bivariate t-copula is:

$$\begin{aligned}
g(u_1, u_2; \rho^*, d) &= \frac{t_{(d, \rho^*)}(t^{-1}(u_1; d), t^{-1}(u_2; d))}{f_{t(d)}(t^{-1}(u_1; d); d) f_{t(d)}(t^{-1}(u_2; d); d)} \\
&= \frac{\Gamma(\frac{d+2}{2})}{\sqrt{1 - \rho^2} \Gamma(\frac{d}{2})} \left(\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d+1}{2})} \right)^2 \\
&\times \frac{\left(1 + \frac{(t^{-1}(u_1; d))^2 + (t^{-1}(u_2; d))^2 - 2\rho t^{-1}(u_1; d)t^{-1}(u_2; d)}{d(1 - \rho^2)} \right)^{-\frac{d+2}{2}}}{\left(1 + \frac{(t^{-1}(u_1; d))^2}{d} \right)^{\frac{d+1}{2}} \left(1 + \frac{(t^{-1}(u_2; d))^2}{d} \right)^{\frac{d+1}{2}}},
\end{aligned} \tag{5.3}$$

where $f_{t(d)}(\cdot; d)$ denotes the PDF of the symmetric univariate t-distribution with d degrees of freedom.

The t-copula will be fitted to the standardized returns of two assets: the S&P 500 and the 10-year Treasury note index (Bloomberg ticker: SPBDU1BT Index). The estimation period spans January 2014 to December 2024, matching the training period used previously for the S&P 500 returns. Initially, univariate models are fitted separately to each asset following the framework outlined in earlier chapters, namely, a volatility model combined with a density model for the standardized asset returns. Because the copula approach requires complete density forecasts as inputs, only the asymmetric t-distribution—among the models discussed in Chapter 3—is considered here. Table 5.1 reports the estimated parameters for the GARCH-Asyt and GJR-Asyt models applied to the Treasury note index.

Model	ω	α	β	θ	d_1	d_2
GARCH-Asyt	1.59e-5	0.202	0.601	-	7.228	0.0279
GJR-Asyt	1.62e-5	0.199	0.595	0.0477	7.224	0.0260

Table 5.1: Bond index estimated parameters.

For each asset, two sets of standardized returns are obtained based on the GARCH(1,1) and GJR(1,1) specifications. This yields four potential copula model combinations. To simplify the analysis, only two cases are considered: i) both assets follow the GARCH-Asyt model, and ii) both follow the GJR-Asyt model. Table 5.2 reports the fitted copula parameters of the GARCH-GARCH and GJR-GJR models.

Copula model	ρ^*	d
GARCH-GARCH	-0.1281	9.0455
GJR-GJR	-0.1249	9.3104

Table 5.2: Copula models estimated parameters.

Using the estimated copula model, it is possible to simulate data and examine whether the model-implied threshold correlations match the empirical pattern observed in Figure 5.2. Figures 5.3 and 5.4 display the simulated threshold correlations obtained from the GARCH-GARCH and GJR-GJR copula models, respectively, alongside the empirical threshold correlations derived from the standardized returns. The t-copula successfully reasonably replicates the increasing correlation observed in the tails. However, it is worth mentioning that the standardized returns exhibit slightly higher correlation in the right tail, a subtle feature the symmetric t-copula cannot fully replicate.

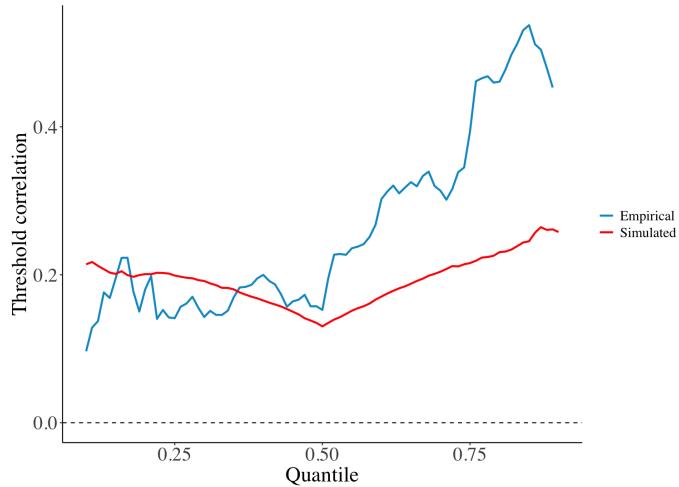


Figure 5.3: Empirical and simulated threshold correlations of GARCH-GARCH copula.

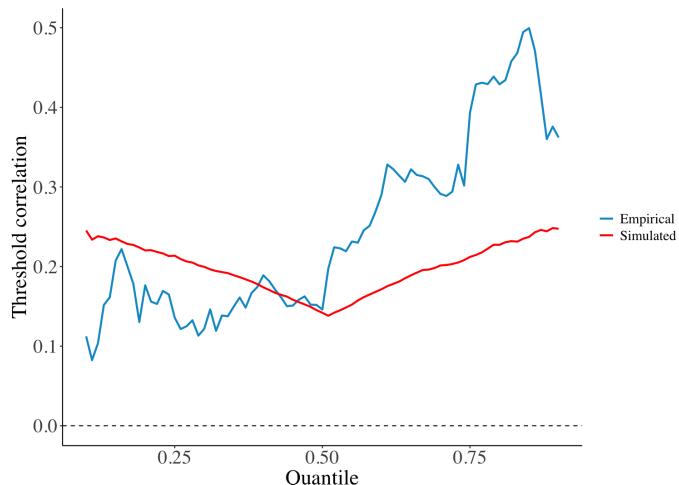


Figure 5.4: Empirical and simulated threshold correlations of GJR-GJR copula.

As shown by Christoffersen and Langlois (2011), dynamic and static asymmetric copulas can match the empirical behavior of threshold correlations and thus offer a powerful alternative to the approach used in this chapter.

The subsequent step is to evaluate the forecasting performance of the copula models. Consider an equally-weighted portfolio composed of the S&P 500 and the bond index. The validity of the multivariate copula models and their effectiveness from a risk management perspective will be assessed using this hypothetical portfolio.

Due to the complexity of the multivariate setup, closed-form expressions for the next day's portfolio density forecast, VaR, and ES are no longer feasible. Consequently, a Monte Carlo approach is used. The algorithm is as follows:

- i. At the end of trading day t , observed returns $R_{1,t}$ and $R_{2,t}$ are used to forecast the next day's conditional volatilities $\sigma_{1,t+1}$ and $\sigma_{2,t+1}$, applying either the fitted GARCH or GJR volatility specification.
- ii. Draw M simulated probability pairs $(\check{u}_{1,t+1}, \check{u}_{2,t+1})$ from the fitted copula model.
- iii. Transform these simulated copula probabilities into standardized shocks using the inverse of the fitted marginal distributions:

$$\check{z}_{i,t+1} = F_i^{-1}(\check{u}_{i,t+1}), \quad i = 1, 2.$$
- iv. Generate asset returns by scaling these shocks with the forecasted volatilities: $\check{R}_{i,t+1} = \sigma_{i,t+1} \check{z}_{i,t+1}, \quad i = 1, 2.$

After simulating M scenarios of next-day returns for each asset, the corresponding portfolio returns are easily computed based on the given portfolio weights. Risk measures, including portfolio VaR and ES, can

then be derived from the simulated distribution of portfolio returns. Specifically, the portfolio VaR at the 2.5% level is estimated as the empirical 2.5th percentile of the simulated portfolio returns. Consequently, at the end of each trading day, the VaR forecast and associated simulated density of the portfolio returns can be obtained.

To further illustrate the simulation procedure, consider the case of a single asset. Figure 5.5 depicts the structure of the Monte Carlo algorithm used to simulate hypothetical next-day returns. Starting from the observed return at time t , the conditional variance forecast σ_{t+1}^2 is obtained using a GARCH or GJR specification. Next, draws are generated from the fitted copula to obtain simulated copula probabilities $\check{u}_{i,t+1}$, which are then transformed into shocks $\check{z}_{i,t+1}$ using the inverse of the marginal CDF. Finally, the shocks are scaled by the forecasted volatility to produce simulated returns $\check{R}_{i,t+1}$.

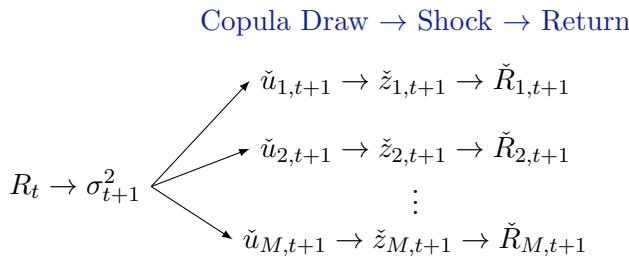


Figure 5.5: Simulation algorithm structure for next day's return.

Once the density forecast and VaR for day $t + 1$ have been generated, the accuracy of the copula-based model can be evaluated. This is done by comparing the realized return R_{t+1} to the simulated distribution, analogous to a Probability Integral Transform (PIT) based now on the Monte Carlo density. The model's effectiveness in risk management is

assessed using VaR validation techniques similar to those applied in Chapter 4.

Figures 5.6 and 5.7 display the 2.5% portfolio VaR along with the breaches observed during the test period. The breaches are almost identical across the two copula models, with the primary weakness again being the lack of independence between violations. Additionally, observe that the GJR-GJR copula model produces sharper increases in VaR following negative returns—reflecting its volatility specification—and thus adapts more quickly, though still not rapidly enough to capture the large negative returns at the onset of President Trump’s tariffs. Table 5.3 summarizes the results of the Christoffersen tests for these models. From a risk management perspective, both copula models may offer an acceptable coverage, but are incapable to adapting quickly to drastic changes in the volatility regime.

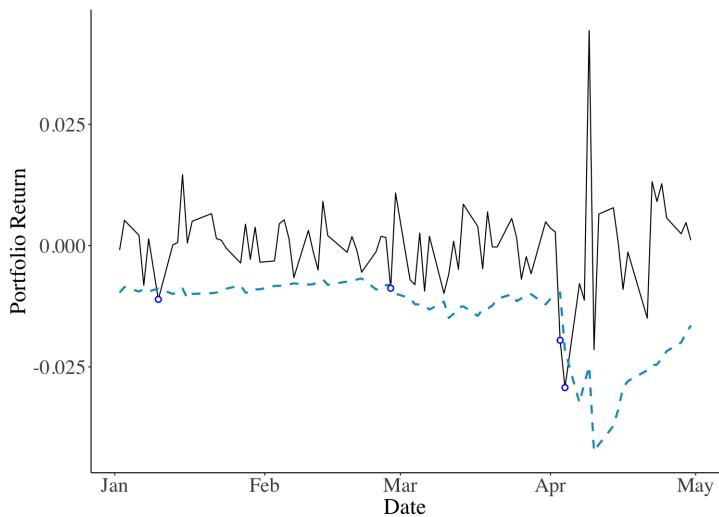


Figure 5.6: 2.5% VaR breaches, GARCH-GACRH Copula.

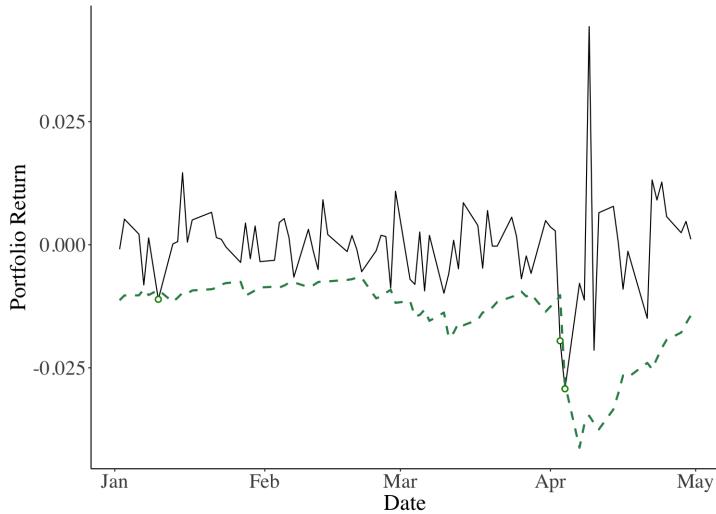


Figure 5.7: 2.5% VaR breaches, GJR-GJR Copula.

Test	p-value	GARCH-GARCH	GJR-GJR
LR_{uc}	Theoretical	0.2138	0.5168
	Simulated	0.166	0.533
LR_{ind}	Theoretical	0.1582	0.0729
	Simulated	0.025	0.012
LR_{cc}	Theoretical	0.1706	0.1622
	Simulated	0.187	0.205

Table 5.3: Theoretical and simulated p -values for VaR testing statistics.

Next, the statistical validity of the copula models is evaluated by examining the portfolio PIT histograms presented in Figures 5.8 and 5.9. No significant deviations from uniformity are apparent. Performing a Kolmogorov–Smirnov test on the normally transformed PIT values,

following the procedure described in [Chapter4](#), yields p -values of 0.5697 for the GARCH-GARCH model and 0.4985 for the GJR-GJR model. These results suggest that the Monte Carlo-generated densities constitute valid density forecasts for the portfolio returns. Figures [5.10](#) and [5.11](#) further illustrate the validity by comparing the empirical CDFs of the normally transformed PIT values against the theoretical standard normal CDF, highlighting the K-S test statistic.

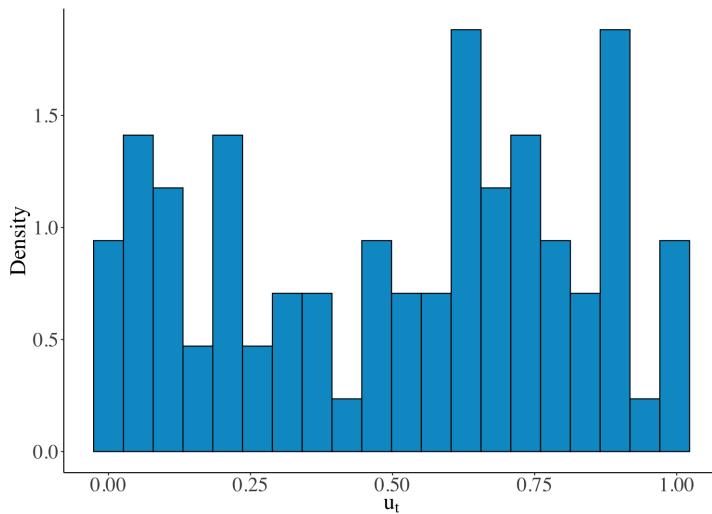


Figure 5.8: Histogram of PIT values, GARCH-GARCH copula.

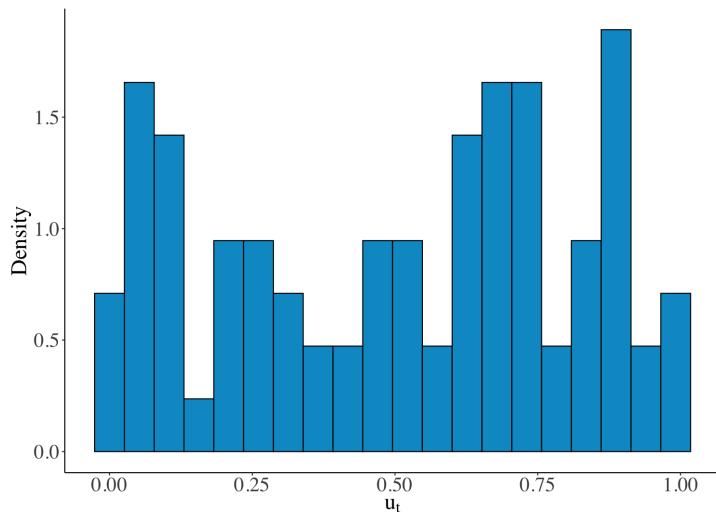


Figure 5.9: Histogram of PIT values, GJR-GJR copula.

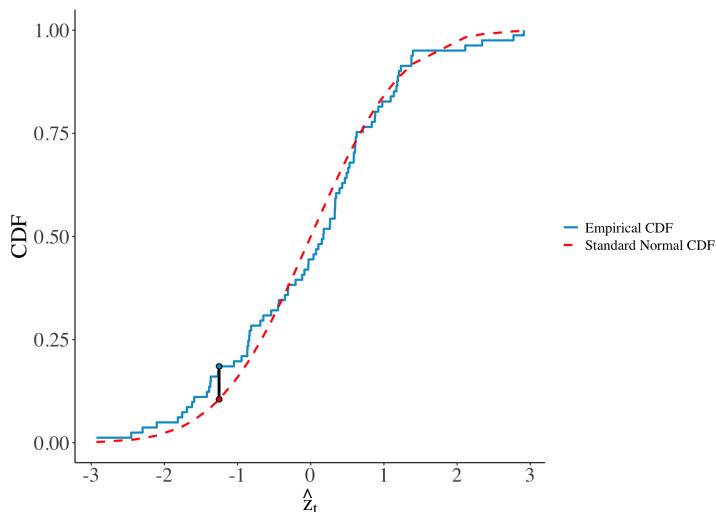


Figure 5.10: Kolmogorov-Smirnov test on \hat{z}_{t+1} , GARCH-GARCH copula

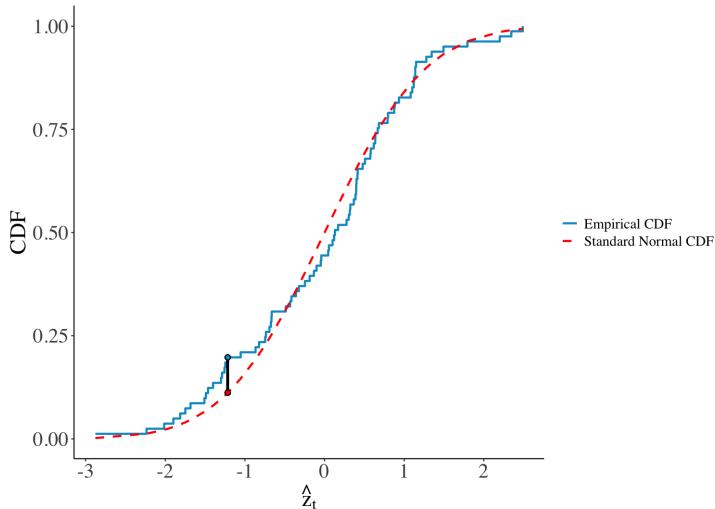


Figure 5.11: Kolmogorov-Smirnov test on \hat{z}_{t+1} , GJR-GJR copula

To conclude, copula models offer two main advantages. First, they allow univariate models for individual asset returns to be estimated independently, leveraging a wide and accessible body of literature. This flexibility makes complex univariate modeling more manageable. Second, copula models effectively capture nonlinear dependencies between asset returns, which are especially important from a risk management perspective. Given the prominence of such nonlinear interactions in financial data, copula models represent a practical and powerful approach to multivariate modeling.

Conclusion

This thesis set out to provide a comprehensive, step-by-step framework for modeling financial returns, beginning with univariate modeling techniques and culminating in a multivariate risk-modeling approach based on copulas. Each chapter built on the previous one, combining theoretical underpinnings with practical implementation for key models in financial econometrics.

[Chapter 1](#) introduced the empirical features of asset returns and framed the modeling challenge. It also described the dataset used throughout: daily returns for the S&P 500 and a 10-year U.S. Treasury note index (Bloomberg ticker: SPBDU1BT Index), with a training sample from January 2015 to December 2024 and a forecasting (test) sample from January 2025 to April 2025.

[Chapter 2](#) addressed the first challenge—modeling dynamic volatility—via GARCH(1,1) and GJR–GARCH(1,1). [Chapter 3](#) turned to the parametric modeling of shocks, considering an asymmetric Student’s t distribution and a peak-over-threshold approach from extreme value theory (EVT).

Chapter 4 evaluated model validity and usefulness for risk management. The chosen GARCH specifications captured the conditional volatility structure, and density-forecast diagnostics suggested statistical adequacy in samples. For risk forecasts, however, models based on the asymmetric t distribution failed to achieve acceptable Value-at-Risk (VaR) coverage and independence of violations, while EVT-based models achieved more acceptable coverage but did not pass independence tests. These results point to hybrid specifications that combine the conservative tail treatment of EVT with shock processes that allow time variation in skewness and kurtosis, thereby adapting more quickly to volatility surges.

Chapter 5 proposed a multivariate framework based on a symmetric t-copula to capture threshold correlations while, for tractability, abstracting from time variation in dependence. In the test set, this static copula struggled with sudden increases in volatility and, consequently, with independence of VaR violations. Nonetheless, PIT-based diagnostics—histograms and Kolmogorov–Smirnov tests on the normal transform—indicated that the copula-based portfolio density forecasts were statistically well calibrated, supporting their validity as density forecasts. To address the remaining limitations, asymmetric and dynamic copulas—potentially combined with time-varying higher moments in the marginal shock processes—are promising extensions.

Throughout the thesis, emphasis was placed on balancing theoretical clarity with empirical rigor. Modeling choices were always accompanied by justification, comparisons with alternatives, and diagnostic tools to evaluate performance. The methodologies were implemented and illustrated using real financial data, with code provided to promote full reproducibility.

Ultimately, the structured modeling path presented here not only equips readers with robust and practical tools to approach financial return modeling, but also underscores the ongoing relevance and vitality of questions first posed decades ago—questions that continue to drive innovation and deepen our understanding of financial markets.

References

- Bai, X., Russell, J. R., & Tiao, G. C. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114(2), 349-360. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304407603000885> doi: [https://doi.org/10.1016/S0304-4076\(03\)00088-5](https://doi.org/10.1016/S0304-4076(03)00088-5)
- Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4), 465–474. Retrieved from <https://doi.org/10.1198/07350010152596718> doi: 10.1198/07350010152596718
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Finance*, 31(3), 307-327. doi: [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3), 542-547. Retrieved from https://public.econ.duke.edu/~boller/Published_Papers/restat_87.pdf
- Bollerslev, T. (2010). Glossary to ARCH (GARCH). In *Volatility and time series econometrics: Essays in honor of robert engle*. Oxford University Press. doi: 10.1093/acprof:oso/9780199549498.001.0001
- Bollerslev, T. (2023). The story of GARCH: A personal odyssey. *Journal of Econometrics*, 234, 96-100. doi: <https://doi.org/10.1016/j.jeconom.2023.01.015>
- Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics*

- of financial markets.* United States: Princeton University Press.
- Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862. Retrieved from <https://doi.org/10.2307/2527341> doi: 10.2307/2527341
- Christoffersen, P. (2012). *Elements of financial risk management.* United States: Academic Press.
- Christoffersen, P., & Langlois, H. (2011). *The joint dynamics of equity market factors* (Working Paper No. CReATES 2011-45). CReATES, Aarhus University; University of Toronto. Retrieved from <https://doi.org/10.1017/S0022109013000598> (Later published in *JFQA* (2013), vol. 48, no. 5, pp.1371–1404) doi: 10.1017/S0022109013000598
- Christoffersen, P., & Pelletier, D. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2(1), 84–108. Retrieved from <https://doi.org/10.1093/jjfinec/nbh004> doi: 10.1093/jjfinec/nbh004
- Coles, S. (2001). *An introduction to statistical modeling of extreme values.* London: Springer-Verlag. doi: 10.1007/978-1-4471-3675-0
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1, 223-236. Retrieved from <https://doi.org/10.1080/713665670>
- Corradi, V., & Swanson, N. R. (2006). Predictive density evaluation. In G. Elliott, C. W. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1, pp. 197–284). Amsterdam: Elsevier. Retrieved from [https://doi.org/10.1016/S1574-0706\(05\)01005-6](https://doi.org/10.1016/S1574-0706(05)01005-6) doi: 10.1016/S1574-0706(05)01005-6

- Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4), 863–883. Retrieved from <https://doi.org/10.2307/2527342> doi: 10.2307/2527342
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. Retrieved from <https://doi.org/10.1080/07350015.1995.10524501> doi: 10.1080/07350015.1995.10524501
- Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, 133(2), 443–477. Retrieved from <https://doi.org/10.1016/j.jeconom.2005.02.005> doi: 10.1016/j.jeconom.2005.02.005
- Elliott, G., & Timmermann, A. (2016). *Economic forecasting*. Princeton, NJ: Princeton University Press. Retrieved from <https://press.princeton.edu/books/hardcover/9780691140131/economic-forecasting> doi: 10.1515/9781400880898
- Fernández, C., & Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441), 359–371. Retrieved from <https://www.jstor.org/stable/2669632>
- Francq, C., & Zakoïan, J. M. (2010). *Garch models: Structure, statistical inference and financial applications*. Chichester: John Wiley & Sons. Retrieved from <https://doi.org/10.1002/9780470670057> doi: 10.1002/9780470670057
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess

- return on stocks. *Journal of Econometrics*, 48(5), 1779-1801. doi: <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x>
- Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705-730. Retrieved from <https://www.jstor.org/stable/2527081>
- Jondeau, E., & Rockinger, M. (2006). The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25(5), 827-853. Retrieved from <https://doi.org/10.1016/j.jimfin.2006.04.007> doi: 10.1016/j.jimfin.2006.04.007
- Mazzeu, J. H. G., Ruiz, E., & Veiga, H. (2018). Uncertainty and density forecasts of arma models: Comparison of asymptotic, bayesian, and bootstrap procedures. *Journal of Economic Surveys*, 32(2), 388–419. Retrieved from <https://doi.org/10.1111/joes.12197> doi: 10.1111/joes.12197
- McLeod, A. I., & Li, W. K. (1983). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4(4), 269–273. doi: <https://doi.org/10.1111/j.1467-9892.1983.tb00373.x>
- McNeil, A. J. (1999). Extreme value theory for risk managers. In *Internal modelling and cad ii* (pp. 93–113). London: Risk Books. Retrieved from <http://www.math.ethz.ch/~mcneil>
- McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4), 271–300. Retrieved from [https://doi.org/10.1016/S0927-5398\(00\)00012-8](https://doi.org/10.1016/S0927-5398(00)00012-8) doi: 10.1016/S0927-5398(00)00012-8

- Okimoto, T. (2008). New evidence of asymmetric dependence structures in international equity markets. *Journal of Financial and Quantitative Analysis*, 43(3), 787–815. Retrieved from <https://doi.org/10.1017/S0022109000004294> doi: 10.1017/S0022109000004294
- Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, 3(1), 15–102. doi: 10.1016/0927-5398(95)00020-8
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2), 527–556. Retrieved from <https://doi.org/10.1111/j.1468-2354.2006.00387.x> doi: 10.1111/j.1468-2354.2006.00387.x
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256. Retrieved from <https://doi.org/10.1016/j.jeconom.2010.03.034> doi: 10.1016/j.jeconom.2010.03.034
- Patton, A. J., & Oh, D.-H. (2011). *Modelling dependence in high dimensions with factor copulas* (Working Paper). Duke University. Retrieved from https://public.econ.duke.edu/~ap172/Oh_Patton_MV_factor_copula_apr12.pdf (Precursor to *JBES* 2017 paper on factor copulas)
- Tsay, R. S. (2010). *Analysis of financial time series*. United States: Wiley.