Modeling Financial Returns with Normal and Non-Normal Distributions: A Tutorial with R Implementation

Miura Financial Research

Contributor to this tutorial is:

Francesco De Simone

Outline

1. Random Variables and Probability Distributions
2. Normal Distribution
3. Fat Tails
4. Student's t Distribution
5. Generalized Error Distribution (GED)
6. Cauchy Distribution
7. Skewed Student-t Distribution
8. Skewed Generalized Error Distribution (SGED)
9. Variance-Gamma Distribution
10. Generalized Hyperbolic Distribution (GHD)
11. Assessing Goodness-of-Fit
12. Conclusions

Introduction

In this tutorial, we explore the theoretical foundations and practical aspects of some of the most common probability distributions used to model financial returns.

We will examine how these distributions behave—particularly in the tails—and discuss their ability to capture extreme events through fat-tailed characteristics.

In addition, we will cover methods for fitting these distributions to real-world financial data and introduce tools for evaluating the goodness of fit, with a focus on both visual diagnostics and statistical measures.

Random Variables

A **random variable** is formally defined as a measurable function:

$$X : \Omega \to (E, \mathcal{E})$$

where $(E, \mathcal{E})$ is a **measurable space**, for a real valued RV, this space coincide with the real one. This allows for random variables taking values in vector spaces, function spaces, or other structured sets.
The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of:

- Ω: the **sample space**, representing all possible outcomes;
- $\mathcal{F}$: a σ-algebra on Ω, defining the collection of **measurable events**;
- $\mathbb{P}$: a **probability measure**, assigning probabilities to events in $\mathcal{F}$, with $\mathbb{P} : \mathcal{F} \to [0, 1]$, and $\mathbb{P}(\Omega) = 1$.

Random variables are typically categorized by the nature of their range:

$$\text{Discrete: } X \in \{x_1, x_2, \dots\}, \quad \text{Continuous: } X \in \mathbb{R}$$

In financial modeling:

- **Continuous random variables** model quantities like asset returns, interest rates, and price paths;
- **Discrete random variables** model binary or countable outcomes such as defaults, credit ratings, or trade counts.

Probability Distributions

Every random variable is associated with a **probability distribution**. It defines the **law** governing the relationship between the values taken by the random variable and the probabilities assigned to those values.
For discrete random variables, the distribution is described by a **probability mass function (PMF)**:

$$P(X = x_i) = p_i, \quad \sum_i p_i = 1$$

Here, each p_i represents the probability that X takes the value x_i.
For continuous random variables, the distribution is given by a **probability density function (PDF)**:

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Probability Distributions

Unlike the PMF, the PDF does not give probabilities directly. Instead, the probability that X falls within an interval $[a, b]$ is given by:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

In fact, for continuous random variables:

$$P(X = x) = 0 \quad \text{for any } x \in \mathbb{R}$$

This is why probabilities are defined over intervals, not at specific points.

Probability Distributions

In both cases, the **cumulative distribution function (CDF)** provides the probability that the variable takes a value less than or equal to a given threshold:

$$F_X(x) = \mathbb{P}(X \leq x)$$

In financial applications, the choice of distribution is crucial. Common examples include:

- **Normal distribution**: suitable for simple models assuming symmetry and light tails;
- **Log-normal distribution**: often used to model asset prices, which cannot be negative;
- **Student-t, Variance Gamma, and other heavy-tailed distributions**: employed to account for excess kurtosis, skewness, and extreme events observed in empirical return data.

Expected Value

The most fundamental statistical property of a random variable is its **expected value** (or mean). For a continuous random variable X with density $f_X(x)$, it is defined as:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

The expected value represents the theoretical long-run average outcome of repeated observations of the variable.

In discrete cases, the expected value is computed as:

$$\mathbb{E}[X] = \sum_i x_i\, P(X = x_i)$$

Variance

The **variance** measures how much a random variable deviates from its expected value. It is the second central moment:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

A higher variance indicates greater dispersion around the mean. Its square root, the **standard deviation**, is often used for interpretability.

In empirical finance, variance is a fundamental measure of risk.

Skewness

Skewness is the third standardized moment. It measures the asymmetry of a distribution around its mean:

$$\text{Skewness} = \mathbb{E}\left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\operatorname{Var}(X)}}\right)^3\right]$$

- Positive skew: long right tail (more extreme high values)
- Negative skew: long left tail (more extreme low values)

In finance, skewness is used to detect asymmetry in return distributions, especially downside risk.

Kurtosis

Kurtosis is the fourth standardized moment. It measures the "tailedness" or propensity of a distribution to produce outliers:

$$\text{Kurtosis} = \mathbb{E}\left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}}\right)^4\right]$$

- High kurtosis: heavy tails, more extreme events than a normal distribution
- Low kurtosis: light tails, fewer outliers

Excess kurtosis is defined as:

$$\text{Excess Kurtosis} = \text{Kurtosis} - 3$$

In finance, high kurtosis reflects tail risk and the possibility of large, unexpected losses or gains.

Modeling Financial Returns

In financial econometrics, asset prices are typically modeled through their returns, which are treated as time-indexed random variables.
Let $\{P_t\}_{t=0}^{T}$ denote the asset price at discrete time points. Returns over one period are defined as:

- **Simple returns:**
$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$
- **Log returns:**
$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

Log returns are often preferred in continuous-time modeling due to their time additivity and analytical tractability.
In both cases, returns are modeled as realizations from a stochastic process $\{R_t\}$, where each R_t is a random variable.

Why Model Returns (Instead of Prices)?

Modeling returns rather than raw prices is a standard practice in financial econometrics for several reasons:

- **Stationarity:** Returns are typically more stationary than price levels, which facilitates statistical inference and forecasting;
- **Comparability:** Returns allow comparison across assets with different price levels and units.

Understanding return dynamics is essential for tasks such as volatility modeling, risk management, and derivative pricing.

Importance of Returns Modeling in Finance

The statistical modeling of asset returns is central to numerous applications in finance:

- **Derivative pricing:**
  - Models like Black-Scholes rely on distributional assumptions for returns.
  - Alternative distributions improve pricing accuracy for options, especially in the presence of skewness or fat tails.
- **Risk measurement:**
  - Measures such as Value-at-Risk (VaR) and Expected Shortfall depend heavily on the assumed return distribution.
  - Misestimating tail behavior can lead to severe underestimation of risk.
- **Portfolio optimization:**
  - Estimating expected returns and covariances is essential for optimal asset allocation.
  - Higher-order moments (skewness, kurtosis) also inform downside risk and investor preferences.
- **Forecasting and trading strategies:**
  - Time-series models for returns help identify market trends and predict volatility.
  - These forecasts drive quantitative trading decisions and hedging strategies.

Accurate return modeling underpins both theoretical developments and practical implementations in quantitative finance.

Why Distributional Assumptions Matter

The choice of the probability distribution for asset returns is crucial in financial modeling, as it affects:

- **Pricing of derivatives**, especially options;
- **Tail risk estimation**, such as Value-at-Risk (VaR) and Expected Shortfall; which determines capital requirements and stress testing;
- **Portfolio optimization**, particularly when **higher moments are included**.

Empirical return distributions often deviate from normality, showing:

- Heavy tails (excess kurtosis);
- Asymmetry (skewness);
- Volatility clustering and time dependence.

Modeling returns with realistic distributions improves the robustness and accuracy of financial decisions.

The Normal Distribution

The **Normal distribution** is a fundamental probability distribution in statistics and finance.

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Its probability density function (PDF) is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Key properties:

- Bell-shaped and symmetric around the mean μ;
- Fully characterized by its mean and variance;
- Skewness = 0 (no asymmetry);
- Kurtosis = 3 (mesokurtic: normal tails);
- **Tails decay exponentially**.

The Normal distribution plays a central role in the classical statistical framework due to the Central Limit Theorem.

Normal Distribution

We recall the density function for a standard normal variable z, and illustrate how the shape of the distribution changes as the standard deviation σ increases:

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$

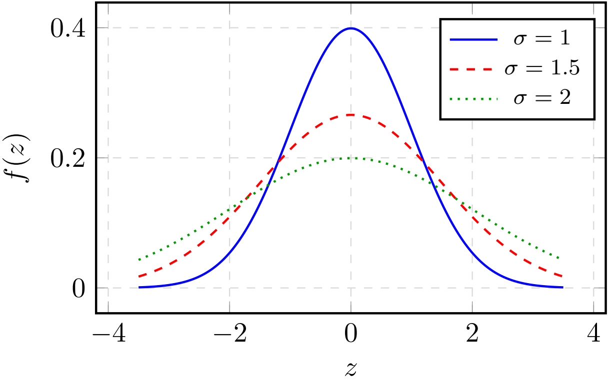

Figure: Comparison of Normal densities with different values of σ.

Applications in Financial Modeling

In classical finance, asset prices $\{P_t\}$ are often modeled as following **Geometric Brownian Motion** (GBM):

$$dP_t = \mu P_t dt + \sigma P_t dW_t$$

Applying Itô's lemma, the log-price evolves as:

$$\log\left(\frac{P_t}{P_0}\right) \sim \mathcal{N}((\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t)$$

This implies:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) \sim \mathcal{N}(\mu, \sigma^2)$$

Interpretation:

- Returns over short time intervals are assumed to be normally distributed;
- This forms the basis for many continuous-time pricing models;
- Log-normality of prices implies normality of log-returns.

Applications in Financial Modeling

The Normal distribution is extensively used in financial theory due to its tractability.

Examples:

- **Black-Scholes-Merton Model:**

$$S_t \sim \text{LogNormal} \Rightarrow r_t \sim \mathcal{N}(\mu, \sigma^2)$$

  Assumes log-returns are normally distributed.
- **Value-at-Risk (VaR):**

$$\text{VaR}_\alpha = \mu + \sigma z_\alpha$$

  where z_α is the standard normal quantile.
- **Portfolio theory (Markowitz):**
  - Assumes returns are normally distributed and characterized fully by mean and covariance.
- **Sharpe Ratio:**

$$\text{SR} = \frac{\mathbb{E}[R] - R_f}{\sigma_R}$$

  Based on normality of returns.

These models offer closed-form solutions and intuitive interpretations.

Estimating μ and σ^2 for a Normal Distribution

When we want to fit a distribution to the data, we need to estimate its parameters. Different approaches are available.
Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution:

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad i = 1, \ldots, n$$

We will derive estimators for μ and σ^2 using:

- Maximum Likelihood Estimation (MLE)
- Method of Moments (MM)

MLE Estimators for μ

The likelihood function is:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2}\right)$$

Taking the log-likelihood:

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

Derive with respect to μ, set to zero:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum (X_i - \mu) = 0 \Rightarrow \hat{\mu}_{\text{MLE}} = \bar{X}$$

MLE Estimator for σ^2

Plug $\hat{\mu} = \bar{X}$ into the log-likelihood and derive w.r.t. σ^2:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum (X_i - \bar{X})^2 = 0$$

Solving:

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Note: This is the *uncorrected sample variance*.

Method of Moments Estimators

From the normal distribution:

$$\mathbb{E}[X] = \mu, \qquad \mathbb{V}[X] = \sigma^2$$

Equating population and sample moments:

$$\hat{\mu}_{\text{MM}} = \frac{1}{n} \sum X_i = \bar{X}$$

$$\hat{\sigma}^2_{\text{MM}} = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

Note: This is the *biased-corrected* (*unbiased*) estimator of the variance.

ML Estimation of μ and σ

R Code

```
# Load library and data
library(quantmod)

# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))
n <- length(x)

# Negative log-likelihood function
nll <- function(params) {
  mu <- params[1]
  sigma <- params[2]
  if (sigma <= 0) return(Inf)
  ll <- -n/2 * log(2*pi) - n/2 * log(sigma^2) -
        1/(2*sigma^2) * sum((x - mu)^2)
  return(-ll)  # Negative for minimization}

# Optimization where the starting parameters are the sample mean and st. dev.
fit <- optim(par = c(mean(x), sd(x)), fn = nll)
# Estimated parameters
mu_hat <- fit$par[1]
sigma_hat <- fit$par[2]
```

Limitations of the Normal Distribution

Despite its widespread use, the normal distribution has significant limitations when applied to empirical return data.

Observed discrepancies in financial returns:

- **Fat tails**;
- **Skewness**: Distributions are often asymmetric;
- **Volatility clustering**: Returns are not i.i.d., violating normality assumptions, more precisely, this lead to a kurtosis > 3;
- **Autocorrelation**: this violates the i.i.d assumptions.

Implications:

- Underestimation of risk in tail events (e.g., financial crises);
- Inaccurate pricing of out-of-the-money options;
- Poor performance in stress testing and backtesting.

Conclusion: More flexible distributions are needed for realistic financial modeling.

Non-Normal Distributions

In financial data analysis, we often encounter the phenomenon of **non-normality** in asset returns. This characteristic is one of the well-documented *stylized facts* of financial markets, and its presence is routinely confirmed through formal statistical tests, such as the Jarque–Bera or Kolmogorov–Smirnov tests. For an extensive discussion, see Campbell et al. [2], and Brooks [1]. Non-normality in this context is typically associated with the presence of **fat tails**, meaning that extreme events (both positive and negative returns) lead to the presence of large losses or gains, than would be predicted under a normal distribution. To account for this, researchers and practitioners often resort to alternative probability distributions that accommodate heavier tails. Common choices include the *Student's t* distribution, the *Cauchy* distribution, *Generalized Error Distribution*, a mixtures or generalizations of these, as well as: Variance Gamma, Generalized Hyperbolic, and skewed versions of these distributions.
This naturally raises the question: *What do we mean by fat tails, and how can we identify them?* In the following slides, we offer a preliminary exploration of this concept. We will examine both their mathematical forms and their graphical representations to better understand how they differ from the normal distribution, particularly in the behavior of their tails.

Fat Tails via Regular Variation

A rigorous definition of fat tails is based on the concept of **regular variation**. A random variable X is said to have fat tails if its distribution function F satisfies:

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\xi}, \quad \forall x > 0, \quad \frac{1}{\xi} > 0.$$

This property characterizes distributions with polynomially decaying tails. The heavier the tails, the smaller the value of $\iota = 1/\xi$, and the slower the tail decays.
In this framework, the tails of the distribution converge to a Fréchet form:

$$H(x) = \exp(-x^{-1/\xi}).$$

Such behavior contrasts with exponentially decaying tails (e.g., Gaussian), and it is typical of distributions used in modeling extreme events and financial returns.

Asymptotic Tail Behavior

If X has regularly varying tails, then for large x, the distribution behaves approximately as:

$$\mathbb{P}(X > x) = 1 - F(x) = Ax^{-1/\xi} + o(x^{-1/\xi}),$$

where $A > 0$ is a scaling constant, and $o(x^{-1/\xi})$ represents higher-order terms vanishing faster than the leading term. This expression implies a **Pareto-type** tail:

$$F(x) \approx 1 - Ax^{-1/\xi}, \quad \text{as } x \to \infty.$$

This means the tail behavior of the distribution is governed by the parameter ξ, which quantifies how "fat" the tails are.

Tail Behavior

Therefore, the tail behavior of a distribution is analyzed through the **survival function**, which we recall is:

$$\bar{F}(x) = \mathbb{P}(X > x)$$

Decay Types:

- **Thin tails (e.g., Gaussian)**:

$$\bar{F}(x) \sim e^{-x^2}, \quad \text{exponential or faster decay}$$

- **Fat tails (e.g., power-law)**:

$$\bar{F}(x) \sim x^{-\iota}, \quad \iota > 0, \text{polynomial decay}$$

We can see from the following Figure, that, the slower the decay of the survival function, the fatter are the tails. We can observe this results also by looking at the distribution function F.

Asymptotic Tail Behavior

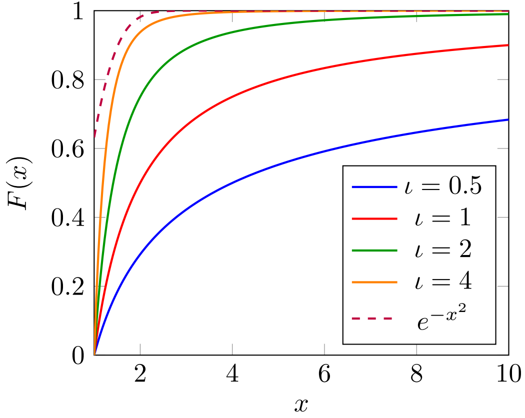

Figure: Comparison of asymptotic tail behavior: power-law tails vs. exponential decay.

Lower values of ι (i.e., higher ξ) correspond to heavier tails. The dashed curve shows the rapid convergence of a Gaussian-type distribution with exponentially decaying tails.

Thin vs. Fat Tails

To sum up what we just mentioned in the previous slides, we classify distributions in:

- **Thin-tailed distributions**: tails decay **exponentially** or faster.
- **Fat-tailed (or heavy-tailed) distributions**: tails decay **slower than exponential**.

Moreover, **Nassim Taleb** distinguishes between:

- **Heavy-tailed distributions**: general class with slower-than-exponential decay.
- **Fat-tailed distributions**: extreme subset where tail events *dominate* the distribution's behavior.

Implication: Fat tails are not just about rare events—they define the distribution's core risk structure.

Thin vs. Fat Tails

Thin-tailed distributions (e.g. Gaussian):

- Exponential decay in the tail
- Effectively bounded: extreme values are negligible
- $\iota \to \infty$
- All moments are finite

Fat-tailed distributions (e.g. Power Laws):

- Polynomial decay in the tail
- Large deviations more likely
- $\iota \in (1, 5)$ often
- Some or all moments may diverge:

$$\begin{cases} \iota < 2 \Rightarrow \text{infinite variance} \\ \iota < 1 \Rightarrow \text{infinite mean} \end{cases}$$

Examples of Fat-Tailed Distributions

1. **Lognormal distribution:**
   - Appears power-law for moderate x
   - But decays faster at extreme values
   - Not a true power-law
2. **Student's t-distribution:**
   - Often used to model financial returns
   - For large degrees of freedom: resembles Gaussian
   - For small ν: true fat tails with finite $\iota = \nu$

Both show heavier tails than the Normal, but only the t-distribution approaches power-law behavior asymptotically.

Scale Invariance in Tail Distributions

A distribution is said to be **scale-invariant** in the tails if a multiplicative change in x results in a proportional change in probability:

$$\mathbb{P}(X > x) = L(x) \cdot x^{-\iota}$$

- $\iota > 0$: **tail index**
- $L(x)$: slowly varying function, such that:

$$\lim_{x \to \infty} \frac{L(cx)}{L(x)} = 1, \quad \text{for any } c > 0$$

This implies a **linear slope in the log-log plot** of the survival function — a hallmark of power laws.

Scale-Invariance vs Scale-Variance

Scale-Invariant (Power Laws):

- Constant rate of decay on log-log scale
- Probability of extreme events decays slowly
- Examples: Pareto, Zipf, Cauchy, some Student-t

Scale-Variant:

- Decay accelerates with x
- Tails curve downward in log-log scale
- Examples: Gaussian, Lognormal

Impact of Tail Index ι

Lower tail indices produce fatter tails and more extreme outcomes:

$$\mathbb{P}(X > x) \sim x^{-\iota}$$

- $\iota < 2 \Rightarrow$ Variance is infinite
- $\iota < 1 \Rightarrow$ Mean is infinite

This does *not* mean tail events are frequent — but when they occur, they are orders of magnitude larger.

Pareto Example

$$\mathbb{P}(X > x) = \left(\frac{x_{\min}}{x}\right)^{\iota}, \quad x \geq x_{\min}$$

Fat Tails and Moments

Another robust and informative method involves analyzing the **raw moments** of a random variable X. The k-th raw moment is defined as:

$$\mathbb{E}[X^k] = \int_0^{+\infty} x^k f_X(x)\,dx = k\int_0^{+\infty} x^{k-1} S_X(x)\,dx,$$

where $f_X(x)$ is the density function and $S_X(x) = \mathbb{P}(X > x)$ is the survival function of X. The existence of these moments depends on the rate at which the survival function decays as $x \to \infty$: the slower the decay, the fewer moments exist.

Fat Tails and Moments

We can define the **maximal finite moment** as:

$$k^* = \sup\left\{k > 0 : \mathbb{E}[X^k] < \infty\right\}.$$

This leads to the following classification:

- If $k^* = \infty$, then X is said to be **light-tailed** based on the moment method;
- If $k^* < \infty$, then X is said to be **heavy-tailed** based on the moment method.

Moreover, for two non-negative random variables X_1 and X_2, with corresponding maximal finite moments k_1^* and k_2^*, we say that X_1 **has a heavier (right) tail than** X_2 if:

$$k_1^* \leq k_2^*.$$

This moment-based approach emphasizes that **fat-tailedness is a property tied to the asymptotic behavior of the distribution**.

Fat Tails and Moments

Despite its intuitive appeal, the moment-based method has several important limitations:

- **Analytical difficulty:** In complex models, computing raw moments can be challenging, making the identification of k^* non-trivial;
- **Theoretical misalignment:** The method does not fully align with classical heavy-tail theory, which typically relies on the existence of the moment generating function (MGF). For instance, the lognormal distribution has all finite moments but no MGF, leading to inconsistent classification;
- **Lack of discrimination among light-tailed distributions:** When comparing two light-tailed distributions with all moments finite, the method cannot distinguish between their tail behaviors.

These issues suggest that while useful, the moment-based method should be applied with care and complemented by other approaches when needed.

Asymptotic Tail Behavior

By exploiting what just mentioned related to the behavior of the survival function we can address the limitations of the moment-based classification, an alternative approach is to compare the **asymptotic behavior of survival functions**.

Let X and Y be two random variables. Define

$$\gamma = \lim_{t \to \infty} \frac{S_X(t)}{S_Y(t)},$$

where $S_X(t) = \mathbb{P}(X > t)$ and $S_Y(t) = \mathbb{P}(Y > t)$ are the survival functions. The tail comparison is as follows:

- X has a **heavier right tail** than Y if $\gamma = \infty$;
- X and Y are **tail-equivalent** if $\gamma = c \in (0, \infty)$;
- X has a **lighter right tail** than Y if $\gamma = 0$.

Fat-Tailed Distributions in Finance

In this next slides, we will explore several probability distributions that are commonly employed to model data with fat tails — that is, distributions that exhibit higher kurtosis. These distributions are particularly useful in fields such as finance, insurance, and risk management, where rare but impactful events play a significant role.
Specifically, we will examine the following distributions: the Student's t-distribution, the Generalized Error Distribution (GED), the Cauchy distribution, the Generalized Hyperbolic Distribution, and the Variance Gamma distribution. Additionally, we will consider skewed versions of some of these distributions, including the skewed Student's t, skewed GED, and skewed Normal distribution, which allow for modeling asymmetry in the data alongside fat tails.

Student's t Distribution

The Student's t-distribution is one of the most widely used fat-tailed distributions, particularly in econometrics and finance due to its ability to capture excess kurtosis and accommodate outliers more effectively than the Normal distribution.

In this presentation, we adopt the parametrization proposed by Bollerslev [3], where the distribution is characterized by three parameters: the location parameter α, the scale parameter β, and the shape parameter ν, which controls the heaviness of the tails.

The probability density function is given by:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\beta\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x-\alpha)^2}{\beta\nu}\right)^{-\frac{\nu+1}{2}} .$$

The distribution is symmetric and unimodal around α, which serves as both the mean and the mode when the mean exists. The tails become heavier as ν decreases. The variance exists only for $\nu > 2$ and is expressed as:

$\mathrm{Var}(x) = \frac{\beta\nu}{\nu-2}$.

As $\nu \to \infty$, the distribution converges to the Normal distribution, while for small values of ν, the tails become significantly heavier, making it suitable for modeling extreme events.

Standardized Student's t Distribution

To facilitate comparison with other distributions and simplify modeling, it is common to standardize the Student's t-distribution by setting its variance to 1. From the variance formula $\text{Var}(x) = \frac{\beta\nu}{\nu-2}$, this standardization implies: $\beta = \frac{\nu-2}{\nu}$. Substituting this into the original density function yields the standardized form:

$$f(z) = \frac{1}{\sigma} \cdot \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu-2)\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{z^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}.$$

Unlike the Normal distribution, where the variable appears in the exponent of a quadratic form, the standardized Student's t features a polynomial decay in the tails. This structure allows the distribution to exhibit *power-law tail behavior*, which becomes more pronounced as the shape parameter ν decreases.
In the upcoming slides, we will visualize how the shape of the distribution changes for different values of ν, clearly illustrating the increasing heaviness of the tails for smaller ν.
Further details on the Student's t and the Generalized Error Distribution can be found in the vignette by Galanos [2].

Student's t Distribution

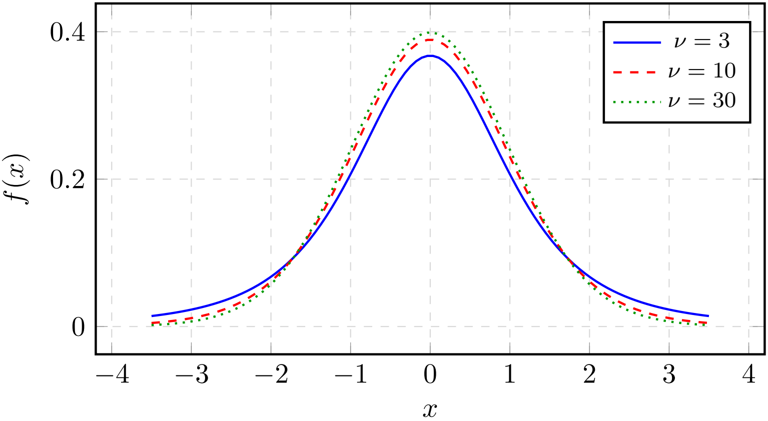

Figure: Comparison of Student-t densities for $\nu = 3, 10$, and 30, illustrating progressively lighter tails.

Estimating μ, σ, and ν via Maximum Likelihood

Given a sample $x_1, x_2, \ldots, x_n$ assumed to follow a Student's t-distribution with parameters μ, σ, and ν, the goal is to estimate these parameters by maximizing the log-likelihood function.

The log-likelihood function is:

$$\mathcal{L}(\mu, \sigma, \nu) = \sum_{i=1}^{n} \log f(x_i \mid \mu, \sigma, \nu)$$

where $f(x_i \mid \mu, \sigma, \nu)$ is the scaled Student's t-density.

We seek:

$$(\hat{\mu}, \hat{\sigma}, \hat{\nu}) = \arg \max_{\mu, \sigma, \nu > 2} \mathcal{L}(\mu, \sigma, \nu)$$

Due to the non-linearity of the function, optimization is performed numerically using bounded methods.

Estimating μ, σ, and ν via Maximum Likelihood

The Student's t-distribution is scaled to ensure unit variance when desired, using the transformation:

$$s = \sqrt{\frac{\nu}{\nu - 2}}, \quad z_i = \frac{x_i - \mu}{\sigma}$$

The custom scaled density is then:

$$f(x_i \mid \mu, \sigma, \nu) = \frac{s}{\sigma} \cdot t_\nu(z_i \cdot s)$$

where $t_\nu(\cdot)$ is the standard Student's t-density with ν degrees of freedom.

The corresponding log-likelihood becomes:

$$\log \mathcal{L}(\mu, \sigma, \nu) = \sum_{i=1}^{n} \log \left(\frac{s}{\sigma} \cdot t_\nu(z_i \cdot s) \right)$$

The optimization routine minimizes the negative log-likelihood:

$$\min_{\mu, \sigma > 0, \nu > 2} \; - \log \mathcal{L}(\mu, \sigma, \nu)$$

Estimating μ, σ, and ν via Maximum Likelihood

We firstly define the standardized Student's t distribution density.

R Code

```
# Standardize Student's t Density
dstd <- function(x, mean = 0, sd = 1, nu = 5, log = FALSE) {
  if (length(mean) == 3) {
    nu <- mean[3]; sd <- mean[2]; mean <- mean[1]
  }s <- sqrt(nu / (nu - 2))
  z <- (x - mean) / sd
  dens <- dt(z * s, df = nu) * s / sd
  if (log) dens <- log(dens)
  dens
}
```

Estimating μ, σ, and ν via Maximum Likelihood

R Code

```
# Define the MLE
stdMLE <- function(x, ...) {
  x <- as.vector(x)
  start <- c(mean = mean(x), sd = sd(x), nu = 5)

  # Negative log-likelihood function
  loglik <- function(par, y) {
    mu <- par[1]; sigma <- par[2]; nu <- par[3]
    if (sigma <= 0 || nu <= 2) return(1e10)

    dens <- dstd(y, mu, sigma, nu)
    if (any(!is.finite(dens)) || any(dens <= 0)) return(1e10)

    -sum(log(dens))
  }# Optimization via nlminb
  fit <- nlminb(start = start,
                objective = loglik,
                lower = c(-Inf, 1e-6, 2.01),
                upper = c(Inf, Inf, Inf),
                y = x, ...)

  names(fit$par) <- c("mean", "sd", "nu")
  return(fit$par)}
```

Estimating μ, σ, and ν via Maximum Likelihood

In the following code we fit the distribution to the S&P 500 returns.

R Code

```
# Load library and data
library(quantmod)
# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))

# Estimate parameters via MLE
params <- stdMLE(x)

print(params)
```

Generalized Error Distribution (GED)

The Generalized Error Distribution (GED), also known as the Exponential Power Distribution, was originally introduced by Nelson [5]. It is widely used in financial econometrics to model the innovation term z_t, especially in the context of GARCH-type models. This distribution is also implemented in the `rugarch` package by Galanos [2].

The probability density function of the GED is defined as:

$$f(x) = \frac{\kappa}{2^{1+\frac{1}{\kappa}}\,\beta\,\Gamma\left(\frac{1}{\kappa}\right)} \exp\left(-\frac{1}{2}\left|\frac{x-\alpha}{\beta}\right|^{\kappa}\right),$$

where α is the location parameter, $\beta > 0$ is the scale parameter, and $\kappa > 0$ is the shape parameter. The distribution is symmetric and unimodal, with α corresponding simultaneously to the mean, median, and mode. When $\kappa = 2$, the GED reduces to the standard Normal distribution. When $\kappa = 1$, it becomes the Laplace (double exponential) distribution. As $\kappa \to \infty$, the density approaches a uniform distribution over a finite support. The shape parameter κ controls the **tail thickness**: smaller values of κ produce heavier tails, allowing for greater kurtosis and making the GED a flexible alternative to the Normal distribution when modeling financial returns.

Standardized Generalized Error Distribution (GED)

To facilitate comparison with other distributions and simplify modeling, we look at standardized version of GED, meaning it has zero mean and unit variance. To obtain a standardized version of the GED, we first compute its theoretical variance:

$$\mathrm{Var}(X) = \beta^2 \cdot 2^{\frac{2}{\kappa}} \cdot \frac{\Gamma\left(\frac{3}{\kappa}\right)}{\Gamma\left(\frac{1}{\kappa}\right)}.$$

Requiring $\mathrm{Var}(X) = 1$, we solve for β and obtain the standardized scale:

$$\beta = \sqrt{2^{-\frac{2}{\kappa}} \cdot \frac{\Gamma\left(\frac{1}{\kappa}\right)}{\Gamma\left(\frac{3}{\kappa}\right)}}.$$

Substituting this into the original density yields the standardized GED:

$$f(z) = \frac{\kappa}{2^{1+\frac{1}{\kappa}}\,\sigma\,\Gamma\left(\frac{1}{\kappa}\right)} \exp\left(-\frac{1}{2}\left|\frac{z}{\sigma}\right|^{\kappa}\right),$$

where σ is the standardized scale parameter:

$$\sigma = \sqrt{2^{-\frac{2}{\kappa}} \cdot \frac{\Gamma\left(\frac{1}{\kappa}\right)}{\Gamma\left(\frac{3}{\kappa}\right)}}.$$

Generalized Error Distribution (GED)

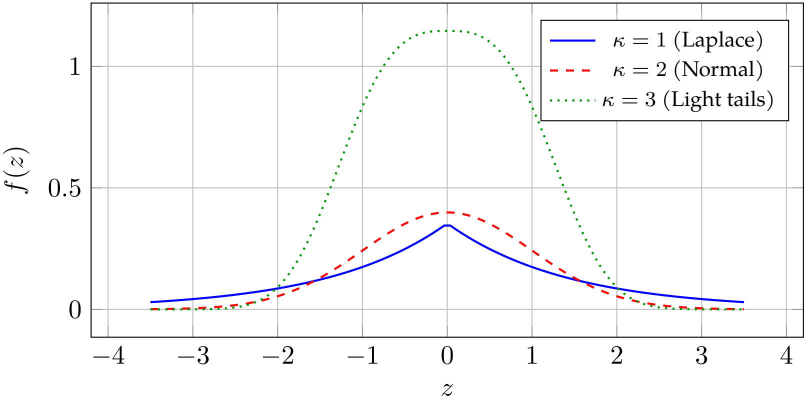

Figure: GED density functions with standardized variance and different values of the shape parameter κ.

ML Estimation of μ σ, and ν

We firstly define the standardized GE distribution density. Here we define μ, σ, and ν, being the location, scale, and shape parameters respectively.

R Code

```
# Standardize GED Density
dged <- function(x, mean = 0, sd = 1, nu = 2, log = FALSE) {
  if (length(mean) == 3) {
    nu <- mean[3]; sd <- mean[2]; mean <- mean[1]
  }
  if (sd <= 0 || nu <= 0) stop("sd and nu must be positive.")

  z <- (x - mean) / sd
  lambda <- sqrt(2^(-2 / nu) * gamma(1 / nu) / gamma(3 / nu))
  g <- nu / (lambda * 2^(1 + 1 / nu) * gamma(1 / nu))
  density <- g * exp(-0.5 * (abs(z / lambda))^nu) / sd

  if (log) log(density) else density
}
```

ML Estimation of μ σ, and ν

R Code

```
# Define the MLE
gedMLE <- function(x) {
  x <- as.vector(x)
  start <- c(mean = mean(x), sd = sd(x), nu = 2)

  neg_loglik <- function(par) {
    mu <- par[1]; sigma <- par[2]; nu <- par[3]
    if (sigma <= 0 || nu <= 0) return(1e10)

    dens <- dged(x, mu, sigma, nu)
    if (any(!is.finite(dens)) || any(dens <= 0)) return(1e10)

    -sum(log(dens))
  }
fit <- optim(start, neg_loglik, method = "L-BFGS-B",
               lower = c(-Inf, 1e-6, 1e-6),
               upper = c(Inf, Inf, Inf))

  names(fit$par) <- c("mean", "sd", "nu")
  return(fit$par)
}
```

ML Estimation of μ σ, and ν

In the following code we fit the distribution to the S&P 500 returns.

R Code

```
# Load library and data
library(quantmod)
# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))

# Estimate parameters via MLE
params <- gedMLE(x)
print(params)
```

Cauchy Distribution

The Cauchy distribution is a continuous, symmetric, and heavy-tailed distribution often used in robust statistics and as a prototype of pathological behavior in probability theory.
Its probability density function is given by:

$$f(x) = \frac{1}{\pi \, \gamma \left[1 + \left(\frac{x - x_0}{\gamma} \right)^2 \right]},$$

where x_0 is the location parameter and $\gamma > 0$ is the scale parameter.
The distribution is centered at x_0, which corresponds to the mode and the median, but not necessarily to the mean. In fact, the Cauchy distribution is notorious for the non-existence of its mean and variance due to the divergence of the corresponding integrals.
This property makes the Cauchy distribution particularly useful for modeling extreme deviations and data with undefined second-order moments. In practical applications, it is also used to test the robustness of estimators under fat-tailed contamination.

Cauchy Distribution

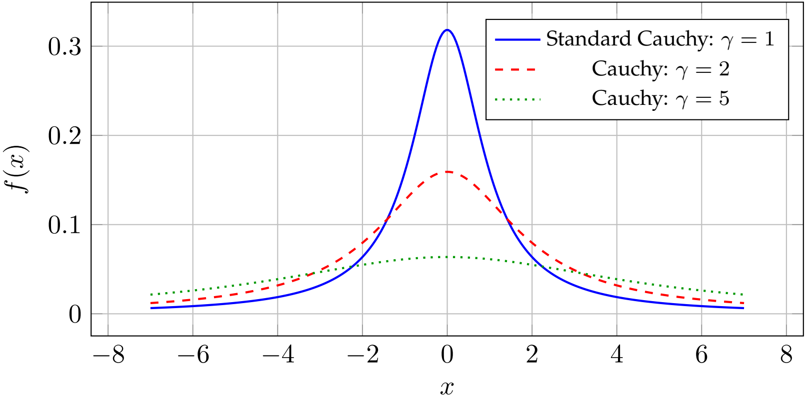

Figure: Cauchy probability density functions with different scale parameters γ.

ML Estimation of x_0 and γ

R Code

```
# Cauchy density (custom wrapper)
dcustom_cauchy <- function(x, location = 0, scale = 1, log = FALSE) {
  dcauchy(x, location, scale, log = log)
}

# Maximum Likelihood Estimation
cauchyFit <- function(x) {
  x <- as.vector(x)
  start <- c(location = median(x), scale = IQR(x) / 2)

  loglik <- function(par) {
    loc <- par[1]
    scale <- par[2]
    if (scale <= 0) return(1e10)
    -sum(dcauchy(x, location = loc, scale = scale, log = TRUE))
  }

  fit <- optim(start, loglik, method = "L-BFGS-B", lower = c(-Inf, 1e-6))
  est <- setNames(fit$par, c("location", "scale"))
  return(est)
}
```

ML Estimation of x_0 and γ

In the following code we fit the distribution to the S&P 500 returns.

R Code

```
# Load library and data
library(quantmod)
# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))

# Estimate parameters via MLE
params <- cauchyFit(x)
print(params)
```

Skewed Student-t Distribution

The Skewed Student-t distribution allows for asymmetry and heavy tails. It introduces a skewness parameter $\xi > 0$ into the standardized Student-t distribution.

We define:

$$m_1 = \frac{2\sqrt{\nu - 2}}{(\nu - 1) \cdot B\left(\frac{1}{2}, \frac{\nu}{2}\right)}, \quad \mu = m_1 \left(\xi - \frac{1}{\xi}\right),$$

$$\sigma = \sqrt{(1 - m_1^2)\left(\xi^2 + \frac{1}{\xi^2}\right) + 2m_1^2 - 1}$$

Let $z = x\sigma + \mu$, and define:

$$\Xi = \xi^{\mathrm{sign}(z)}, \quad g = \frac{2}{\xi + 1/\xi}$$

The standardized density becomes:

$$f(x) = \sigma \cdot g \cdot f_t\left(\frac{z}{\Xi}\right)$$

where $f_t(\cdot)$ is the Student-t density with ν degrees of freedom.

Skewed Student-t Distribution

Parameters:

$\mu \in \mathbb{R}$	location
$\sigma > 0$	scale (standardized to ensure unit variance)
$\nu > 2$	degrees of freedom, controls tail thickness
$\xi > 0$	skewness parameter: asymmetry of the distribution

Special cases:

- If $\xi = 1$, the distribution reduces to a symmetric Student-t.
- If $\xi > 1$, the distribution is right-skewed.
- If $\xi < 1$, the distribution is left-skewed.

This parametrization ensures $\mathbb{E}[X] = 0$ and $\mathbb{V}[X] = 1$, and is implemented in packages like `rugarch` and `fGarch` in R.

Skewed Student-t Distribution

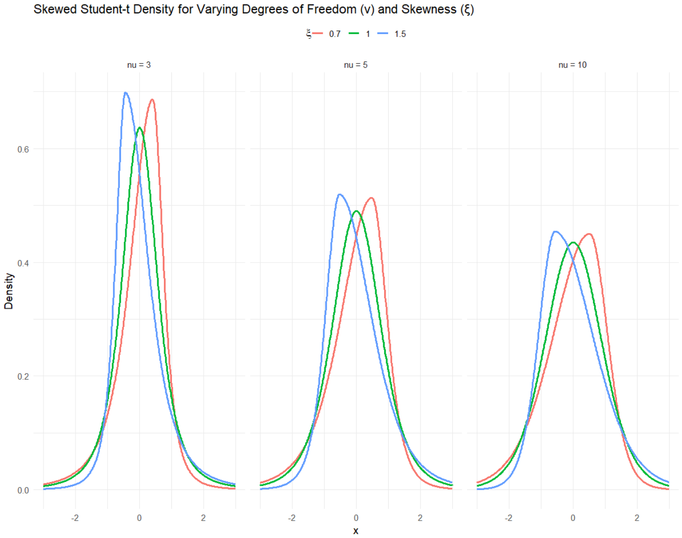

Figure: Skewed Student's t for different ξ and ν parameters.

ML Estimation of μ, σ, ξ and ν

R Code

```
# Skewed Standardized density 1
.dsstd <- function(x, nu, xi) {
  m1 <- 2 * sqrt(nu - 2) / (nu - 1) / beta(1/2, nu/2)
  mu <- m1 * (xi - 1/xi)
  sigma <- sqrt((1 - m1^2) * (xi^2 + 1/xi^2) + 2 * m1^2 - 1)
  z <- x * sigma + mu

  Xi <- xi^sign(z)
  g <- 2 / (xi + 1/xi)
  dens <- g * dstd(z / Xi, nu = nu)
  dens * sigma
}

# Skewed Standardized density 2
dsstd <- function(x, mean = 0, sd = 1, nu = 5, xi = 1.5, log = FALSE) {
  if (length(mean) == 4) {
    xi <- mean[4]; nu <- mean[3]; sd <- mean[2]; mean <- mean[1]
  }
  dens <- .dsstd((x - mean) / sd, nu, xi) / sd
  if (log) dens <- log(dens)
  dens
}
```

ML Estimation of μ, σ, ξ and ν

R Code

```
sstdMLE <- function(x, trace = FALSE, ...) {
  x <- as.vector(x)
  start <- c(mean = median(x), sd = IQR(x) / 2, nu = 5, xi = 1.5)
negloglik <- function(par, y, trace) {
    ll <- tryCatch(-sum(log(dsstd(y, par[1], par[2], par[3], par[4]))),
                 error = function(e) 1e9)
    if (!is.finite(ll)) ll <- 1e9
    if (trace) cat("\nLogLik:", -ll, "\nParams:", par, "\n")
    ll
  }fit <- nlminb(
    start = start,
    objective = negloglik,
    lower = c(-Inf, 1e-6, 2.01, 0.01),
    upper = c(Inf, Inf, 100, 10),
    y = x,
    trace = trace,
    ...
  )names(fit$par) <- c("mean", "sd", "nu", "xi")
  list(
    estimate = fit$par,
    logLik = -fit$objective,
    convergence = fit$convergence,
    method = "MLE")}
```

ML Estimation of μ, σ, ξ and ν

In the following code we fit the distribution to the S&P 500 returns.

R Code

```
# Load library and data
library(quantmod)
# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))

# Estimate parameters via MLE
params <- sstdMLE(x)
print(params)
```

Skewed Generalized Error Distribution (SGED)

The Skewed Generalized Error Distribution (SGED) generalizes the GED by introducing a skewness parameter $\xi > 0$, allowing asymmetric modeling of data.

Let $X \sim \text{SGED}(\mu, \sigma, \nu, \xi)$, with:

- μ: location parameter (mean-like),
- $\sigma > 0$: scale parameter (std deviation-like),
- $\nu > 0$: shape parameter (tail thickness),
- $\xi > 0$: skewness parameter.

The density is defined as:

$$f_X(x) = \frac{1}{\sigma} \cdot f_Z\left(\frac{x-\mu}{\sigma}\right)$$

where Z is the standardized SGED.

Standardized Skewed GED

The standardized SGED density $f_Z(z)$ is:

$$f_Z(z) = \sigma' \cdot g(z \cdot \sigma' + \mu')$$

where:

$$g(u) = \frac{2}{\xi + \frac{1}{\xi}} \cdot f_{\text{GED}}\left(\frac{u}{\xi^{\text{sign}(u)}}\right)$$

The kernel function $f_{\text{GED}}(x)$ is the Generalized Error Distribution:

$$f_{\text{GED}}(x) = \frac{\nu}{2\beta\Gamma(1/\nu)} \exp\left(-\left|\frac{x}{\beta}\right|^{\nu}\right)$$

This density is symmetric and controlled by the shape parameter ν.

Standardized Skewed GED

The constants used in the standardized SGED density are:

$$\beta = \sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}, \quad \lambda = \sqrt{2^{-2/\nu} \cdot \frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}}$$

Skew-normalization moments:

$$m_1 = 2^{1/\nu} \lambda \cdot \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)}$$

Standardization shift and scale:

$$\mu' = m_1 \left(\xi - \frac{1}{\xi} \right), \quad \sigma' = \sqrt{(1 - m_1^2)(\xi^2 + \frac{1}{\xi^2}) + 2m_1^2 - 1}$$

These ensure that Z is approximately standardized:

$$\mathbb{E}[Z] \approx 0, \quad \mathrm{Var}(Z) \approx 1$$

Skewed GED Distribution

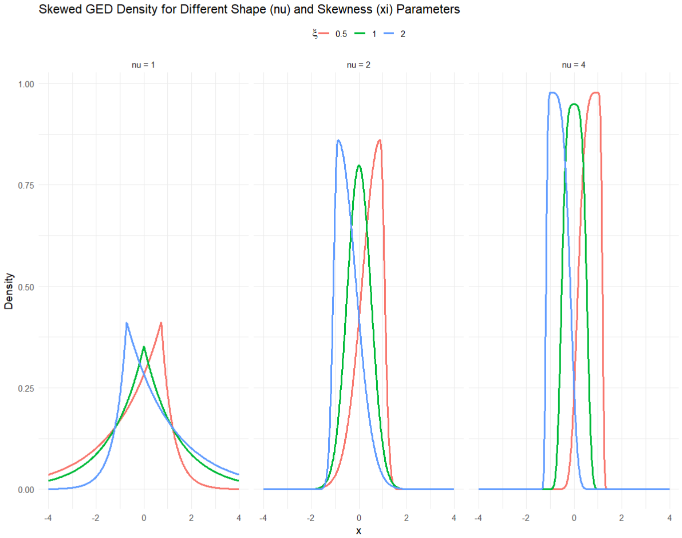

Figure: Skewed GED for different ξ and ν parameters.

ML Estimation of μ, σ, ξ and ν

R Code

```
# SGED Density 1
.dsged <- function(x, nu, xi) {
  lambda <- sqrt(2^(-2/nu) * gamma(1/nu) / gamma(3/nu))
  m1 <- 2^(1/nu) * lambda * gamma(2/nu) / gamma(1/nu)
  mu <- m1 * (xi - 1/xi)
  sigma <- sqrt((1 - m1^2) * (xi^2 + 1/xi^2) + 2 * m1^2 - 1)

  z <- x * sigma + mu
  Xi <- xi^sign(z)
  g <- 2 / (xi + 1/xi)
  dens <- g * dged(z / Xi, nu)
  dens * sigma
}

# SGED Density 2
dsged <- function(x, mean = 0, sd = 1, nu = 2, xi = 1.5, log = FALSE) {
  if (length(mean) == 4) {
    xi <- mean[4]; nu <- mean[3]; sd <- mean[2]; mean <- mean[1]
  }
  if (sd <= 0 || nu <= 0 || xi <= 0) stop("sd, nu and xi must be > 0.")
  dens <- .dsged((x - mean)/sd, nu, xi) / sd
  if (log) log(dens) else dens}
```

ML Estimation of μ, σ, ξ and ν

R Code

```
# MLE fitting
sgedMLE <- function(x, trace = FALSE, ...) {
  x <- as.vector(x)
  start <- c(mean = mean(x), sd = sd(x), nu = 2, xi = 1.5)

  loglik <- function(par) {
    if (any(par[c("sd", "nu", "xi")] <= 0)) return(1e10)
    -sum(log(dsged(x, par["mean"], par["sd"], par["nu"], par["xi"])))
  }

  fit <- nlminb(start = start, objective = loglik,
                lower = c(-Inf, 1e-6, 1e-6, 1e-6),
                upper = rep(Inf, 4))
  names(fit$par) <- c("mean", "sd", "nu", "xi")

  fit$par
}
```

ML Estimation of μ, σ, ξ and ν

In the following code we fit the distribution to the S&P 500 returns.

R Code

```
# Load library and data
library(quantmod)
# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))
# Estimate parameters via MLE
params <- sgedMLE(x)
print(params)
```

Variance-Gamma Distribution

The VG formulation and name come from the idea of using a normal distribution with a mean of zero and a variance that has a gamma distribution. The VG distribution, is very popular within the option pricing context, indeed, it is particularly useful for modeling asset returns, since, it might account for jumps.
Let $X \sim VG(\gamma, \sigma, \theta, \nu)$. The density function is given by:

$$f_X(x) = \frac{2\exp\left(\frac{\theta(x-\gamma)}{\sigma^2}\right)}{\nu^{1/\nu}\sqrt{2\pi}\,\sigma\,\Gamma(1/\nu)} \cdot \left(\frac{|x-\gamma|}{\sqrt{2\sigma^2/\nu+\theta^2}}\right)^{1/\nu-1/2} \cdot K_{1/\nu-1/2}\left(\frac{|x-\gamma|}{\sigma^2}\sqrt{2\sigma^2/\nu+\theta^2}\right)$$

where $K_\lambda(\cdot)$ is the modified Bessel function of the second kind.
The parameters of the VG distribution are:

$\gamma \rightarrow$ location (VG constant)

$\sigma \rightarrow$ scale parameter

$\theta \rightarrow$ asymmetry (skewness)

Variance-Gamma Distribution

- For small ν, the distribution has heavier tails.
- Positive θ induces right skewness; negative θ induces left skewness.
- When $\theta = 0$, the distribution is symmetric.

Moreover,

- The density is computed using the modified Bessel function $K_\lambda(z)$.
- Near $x = \gamma$, special care is required to avoid numerical instability.
- The density may return ∞ when $x = \gamma$ for certain values of ν.

Variance-Gamma Distribution

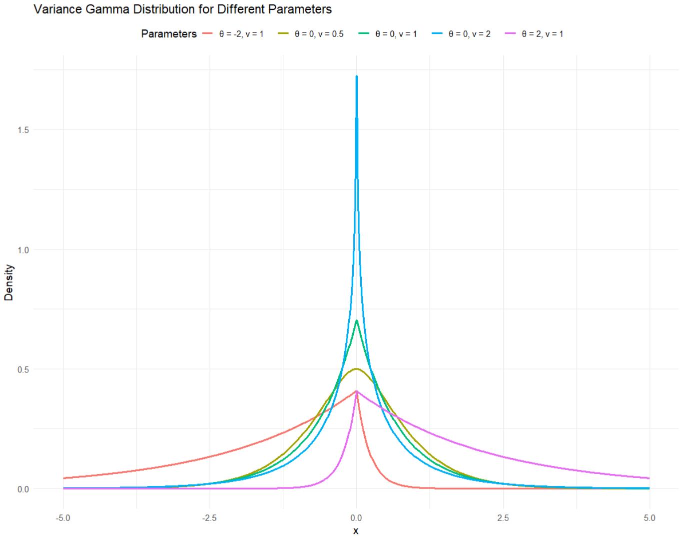

Figure: Variance Gamma Distribution for different θ and ν parameters.

ML Estimation of VG Parameters

In the following code we fit the distribution to the S&P 500 returns. We use the `VarianceGamma` package present in R for the fitting process.

R Code

```
# Load library and data
library(quantmod)
library(VarianceGamma)
# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))
# Estimate parameters via MLE
params <- vgFit(x=x)
print(params)
```

Generalized Hyperbolic Distribution (GHD)

Let $X \sim GH(\lambda, \alpha, \beta, \delta, \mu)$. Its probability density function is given by:

$$f_X(x) = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\,\delta^\lambda\,K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} \cdot \frac{K_{\lambda - \frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{\frac{1}{2} - \lambda}} \cdot \exp\left(\beta(x - \mu)\right)$$

where $K_\lambda(\cdot)$ is the modified Bessel function of the second kind.

The parameters are interpreted as:

$$\begin{aligned}
\lambda &\to \text{shape parameter controlling tail behavior} \\
\alpha &\to \text{tail heaviness / steepness} \\
\beta &\to \text{skewness parameter} \\
\delta &\to \text{scale parameter} \\
\mu &\to \text{location parameter}
\end{aligned}$$

Generalized Hyperbolic Distribution (GHD)

- The GHD includes many important special cases, such as the Normal Inverse Gaussian (NIG) when $\lambda = -\frac{1}{2}$.
- The skewness parameter β controls asymmetry: positive β skews right, negative skews left.
- The parameter α controls tail thickness: larger α corresponds to lighter tails.
- The shape parameter λ adjusts tail decay and kurtosis.
- Numerical evaluation of the density requires stable computation of modified Bessel functions, especially near the mode.
- For $\delta \to 0$, the distribution converges to a degenerate distribution centered at μ.

Generalized Hyperbolic Distribution (GHD)

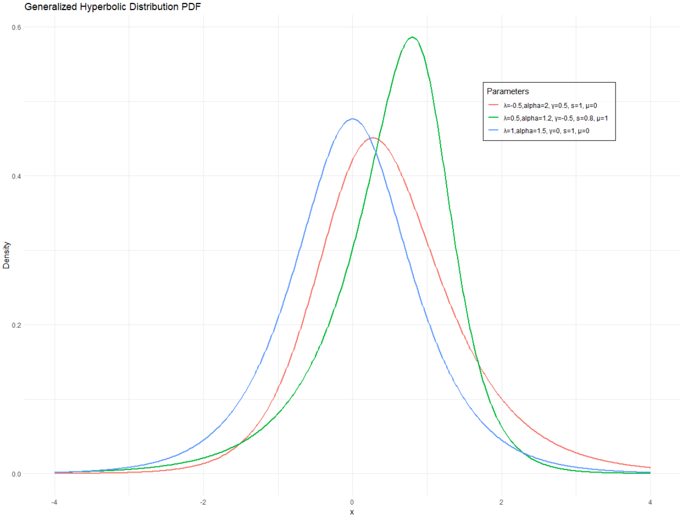

Figure: Generalized Hyperbolic Distribution for different parameters.

ML Estimation of GHD Parameters

In the following code we fit the distribution to the S&P 500 returns. We use the ghyp package present in R for the fitting process.

R Code

```
# Load library and data
library(quantmod)
library(ghyp)
# Download data from Yahoo Finance of S&P 500
getSymbols("^GSPC", from = "2015-01-01", to = "2023-12-31")
log_ret <- dailyReturn(Cl(GSPC), type = "log")
x <- as.numeric(na.omit(log_ret))
# Estimate parameters via MLE
params <- fit.ghypuv(data=x)
print(params)
```

Are these Distribution Fat-Tailed?

How do we know, if the distributions just considered are fat tailed? After introducing the concept of fat tails, a key insight comes from analyzing the CDF and its asymptotic behavior.

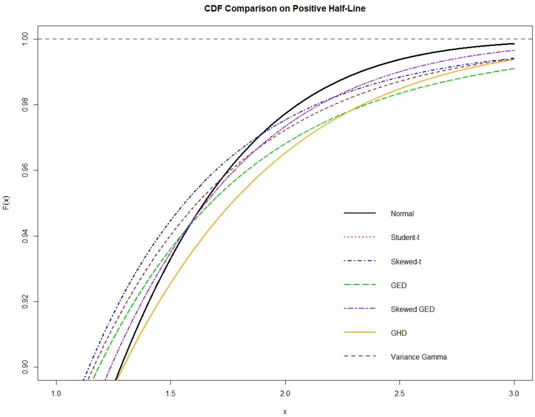

Figure: CDF for Normal, Student's t, GED, Skewed Student-t, and Skewed GED, GHD, and Variance Gamma.

Are these Distributions Fat-Tailed?

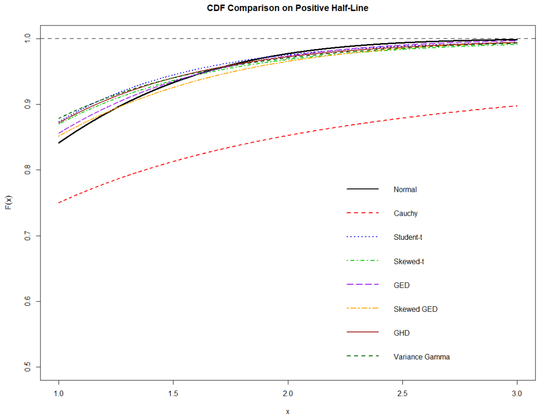

Figure: CDF for Normal, Student's t, GED, Skewed Student-t, and Skewed GED, GHD, Variance Gamma, and Cauchy

Are these Distributions Fat-Tailed?

From the previous slides, we observe that all distributions—except for the Normal—exhibit slower convergence of their cumulative distribution functions to 1, indicating the presence of fat tails.

In this slide, we exclude the Cauchy distribution from the comparison, as its tails are significantly heavier than all others. The Cauchy distribution's extreme tail behavior is so dominant that it obscures meaningful differences among the other distributions when plotted together.

These observations underscore how tail behavior varies across different models and emphasize the importance of selecting appropriate distributions for modeling financial returns, where extreme events and heavy tails are common.

This highlights why understanding and accurately modeling tail behavior is critical in financial risk management and asset pricing.

Mixture Distributions

A **mixture distribution** models a probability distribution as a weighted combination of two or more component distributions.

Formally, if $X \sim \sum_{i=1}^{K} \pi_i f_i(x)$, then:

- π_i: weight of component i, with $\sum \pi_i = 1$, $\pi_i \geq 0$
- $f_i(x)$: density of the i-th component distribution
- K: number of components

Example:

$$f_X(x) = \pi \cdot f_{\text{Normal}}(x) + (1 - \pi) \cdot f_{\text{Cauchy}}(x)$$

Why Mixture Models in Finance?

Mixture models are widely used in finance due to the complex and heterogeneous nature of financial data. Key motivations include:

- **Modeling heavy tails:** Asset returns often exhibit fat tails that cannot be captured by a single Gaussian distribution.
- **Capturing volatility regimes:** Returns may alternate between low- and high-volatility periods (e.g., during crises).
- **Skewness and asymmetry:** Financial return distributions are frequently asymmetric; mixtures help model such skewness.
- **Multimodal behavior:** In portfolio returns or derivative pricing, returns may come from distinct subpopulations.
- **Improved risk management:** More accurate modeling of extreme events and tail risk, e.g., for Value-at-Risk (VaR) estimation.
- **Flexible density estimation:** Enables non-parametric-like flexibility while maintaining parametric interpretability.

Assessing Goodness-of-Fit

To evaluate whether a proposed probability distribution fits observed data, we use statistical goodness-of-fit tests.

- These tests compare the observed data with the theoretical distribution.
- Null hypothesis: the data follow the specified distribution.
- Common tests include:
  - Chi-squared test
  - Kolmogorov–Smirnov (KS) test
  - Anderson–Darling test (less common, more powerful)

Chi-Squared Goodness-of-Fit Test

Chi-squared test compares observed and expected frequencies in bins.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

- O_i: observed frequency in bin i
- E_i: expected frequency under the assumed distribution
- k: number of bins

Notes:

- Requires binning of continuous data.
- Expected counts E_i should be 5 for validity.

Kolmogorov–Smirnov (KS) Test

The KS test compares the empirical distribution function (EDF) with the cumulative distribution function (CDF) of a reference distribution.

$$D = \sup_x |F_n(x) - F(x)|$$

- $F_n(x)$: empirical CDF of the sample
- $F(x)$: theoretical CDF
- D: maximum absolute difference

Notes:

- Sensitive to deviations in location, scale, or shape.

Model Assessment in Practice

- Use histograms and density plots to visualize fit.
- Apply statistical tests (e.g., KS or Chi-squared).
- Consider the p-value:
  - If $p < 0.05$, reject the null hypothesis (the distribution is not a good fit).
  - If $p \geq 0.05$, we do not reject the fit.
- Combine visual and statistical tools for robust assessment.

Normal vs. Non-Normal distributions

The figure below illustrates the distributions just mentioned, fitted to the returns of the S&P 500.

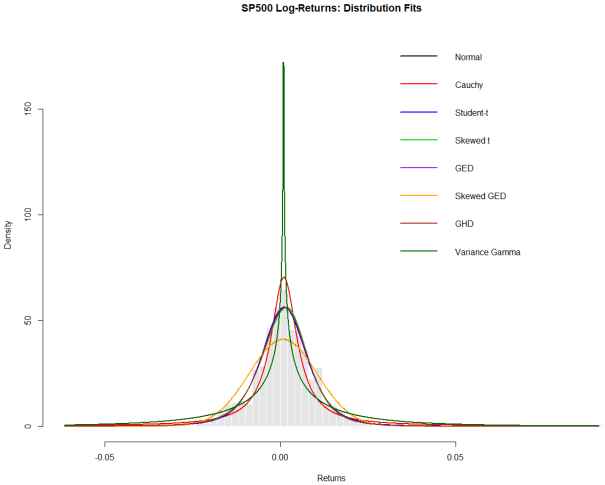

Figure: S&P 500 and fitted distributions.

Normal vs. Non-Normal distributions

The previous slides reveal a clear departure from the Normal distribution. Visually, both the Cauchy and Student's t distributions appear to provide a superior fit to the data. Nevertheless, a formal statistical test is necessary to evaluate adherence to the assumed null distributions.

Conclusions

- **Random variables** are fundamental tools in financial modeling. They can be discrete or continuous depending on the nature of the problem (e.g., default events vs. asset returns).
- The **normal distribution** is widely used in finance, especially due to its role in classical models such as Black-Scholes for option pricing and standard risk management frameworks.
- **Fat tails** are identified through the *asymptotic behavior* of the distribution. A power-law decay implies heavy (fat) tails, while exponential decay—as in the normal distribution—implies thin tails.
- Modeling fat tails is essential in finance to capture the higher probability of extreme events. Common distributions used include: Student's t, Generalized Error Distribution (GED), their skewed versions, Cauchy, Variance Gamma, and Generalized Hyperbolic.

References (1/2)

Brooks, C. (2019). *Introductory Econometrics for Finance* (4th ed.). Cambridge University Press.

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., & Whitelaw, R. F. (1998). *The Econometrics of Financial Markets*. Princeton University Press.

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.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of Econometrics* (Vol. 4, pp. 2959–3038). Elsevier.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), 347–370.

References (2/2)

Boshnakov, G. N., Wuertz, D., Chalabi, Y., Setz, T., & Maechler, M. (2025). fGARCH: Autoregressive Conditional Heteroskedastic Modelling (Version 4033.92) [Software per computer].Retrieved from `https://cran.r-project.org/web/packages/fGarch/index.html`

Galanos, A. (2024). *rugarch: Univariate GARCH Models* (R package version 1.5-3) [Computer software]. Retrieved from `https://cran.r-project.org/web/packages/rugarch/index.html`