

Forecasting VaR using GARCH and KDE

Leonardo Cruciani

April 10, 2025

1 Introduction

Modeling and forecasting risk is a cornerstone of financial econometrics. Among risk measures, Value-at-Risk (VaR) is one of the most widely used tools for quantifying potential losses in portfolios over a given time horizon and confidence level, as discussed in Jorion (2006). However, the accuracy of VaR depends critically on how well the distribution of returns is modeled, especially in the tails. Financial returns typically exhibit heavy tails, skewness, and volatility clustering—features that violate the assumptions of normality underlying classical VaR approaches like RiskMetrics. To address these stylized facts, the paper by Lehnert and Wolff (2004) proposes a *scale-consistent* VaR framework that combines truncated skewed Lévy distributions (TSL) with an augmented GARCH volatility modeling (Pagan and Schwert 1990). This approach captures both the time-varying volatility and non-Gaussian behavior of asset returns and improves longer horizon VaR forecasting through a scaling rule derived from Lévy flight dynamics. In this paper, I first (sections 3 and 4) replicate the method of Lehnert and Wolff (2004) to forecast Value-at-Risk using GARCH-filtered returns modeled with a truncated skewed Lévy distribution. Here, I use Kernel Density Estimation (KDE) (Silverman 1986) to improve the computational complexity of the task of calibrating the TSL shape parameters. The accuracy of the VaR estimates is evaluated by means of a backtesting procedure based on the number of breaches, following the methodology proposed in Christoffersen (2003). Then (in section 5), motivated by the rigidity of parametric approaches, I introduce a nonparametric alternative based on empirical quantiles and drop the TSL distribution assumption. This nonparametric approach offers both advantages and disadvantages: it requires fewer distributional assumptions, is relatively easy to implement and offers overall better results, but at the expense of requiring more data for empirical quantiles to accurately estimate their true values.

2 Data

The dataset contains the closing prices of the S&P500 index from 2006 to 2024. The dataset is split in two, with the first 11 years being used for training and the last 7 years for testing. From the dataset the log returns are computed, expressed in percentage terms, and aggregated for 5, 10 and 20 days.

3 GARCH

The paper Lehnert and Wolff (2004) uses an augmented GARCH model that nests some of the most commonly used (e.g. GARCH(1,1), GJR-GARCH, Aparch, etc.) generalizations. In this work I

Dep. Variable:	Returns	R-squared:	0.000
Mean Model:	Constant Mean	Adj. R-squared:	0.000
Vol Model:	Asym. Power GARCH (power: 1.8)	Log-Likelihood:	-3683.10
Distribution:	Normal	AIC:	7376.21
Method:	Maximum Likelihood	BIC:	7405.84
		No. Observations:	2769
Date:	Tue, Apr 08 2025	Df Residuals:	2768
Time:	17:06:58	Df Model:	1

	coef	std err	t	P> t	95.0% Conf. Int.
mu	0.0263	1.466e-02	1.792	7.309e-02	[-2.458e-03,5.501e-02]
	coef	std err	t	P> t	95.0% Conf. Int.
ω	0.0247	5.356e-03	4.605	4.124e-06	[1.417e-02,3.516e-02]
γ	0.0000	2.057e-02	0.000	1.000	[-4.032e-02,4.032e-02]
η	0.2059	3.003e-02	6.858	7.000e-12	[0.147, 0.265]
δ	0.8818	1.835e-02	48.043	0.000	[0.846, 0.918]

Table 1: Constant Mean - Asym. Power GARCH (power: 1.8) Model Results. Covariance estimator: robust

will resort to a simpler Power GARCH model with leverage effect to model the volatility σ_t :

$$\sigma_t^\kappa = \omega + \gamma |\epsilon_{t-1}|^\kappa + \delta \sigma_{t-1}^\kappa + \eta |\epsilon_{t-1}|^\kappa I_{[\epsilon_{t-1} < 0]} \quad (1)$$

Where $\epsilon_t \sim D(0, 1)$ are independent shocks distributed according to a *standardized distribution* $D(0, 1)$. In the paper Lehnert and Wolff (2004) and in section 4 this $D(0, 1)$ distribution is the TSL calibrated on historical data, while in section 5 the KDE distribution is used. The parameters to be estimated are: γ , which models the sensitivity of the volatility to the shock; δ , which models the persistence of the volatility; η , which models the leverage effect; the exponent κ (typically $\simeq 2$) and the so called (when $\kappa = 2$) long term variance ω .

To model the return r_t I used the following model:

$$r_t = \mu + \sigma_t \epsilon_t, \quad (2)$$

Where the mean μ can be safely assumed to be constant for medium to short term returns. The parameters of the GARCH are calibrated on historical data and the model is then used to forecast the next period variance $\hat{\sigma}_{t+1}^2$ as returns become available. Therefore, the closing prices in the dataset are standardized following the formula:

$$z_t = \frac{r_t - \hat{\mu}}{\hat{\sigma}_t} \quad (3)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are obtained from the GARCH.

4 Paper Replication

4.1 Truncated Skewed Levy

The $TSL(\mu, c, \alpha, \lambda, \beta)$ is defined through its characteristic function $\varphi_{TSL}(k) = \exp(-\psi(k; \mu, c, \alpha, \lambda, \beta))$ where:

$$\begin{aligned} \psi(k; \mu, c, \alpha, \lambda, \beta) = & ik\mu + \\ & + c^\alpha \left[\frac{\lambda^2 - (k^2 + \lambda^2)^{\alpha/2}}{\cos(\pi\alpha/2)} \cos\left(\alpha \operatorname{atan}\left(\frac{|k|}{\lambda}\right)\right) \right] \times \\ & \times \left[1 + i \operatorname{sgn}(k) \beta \tan\left(\alpha \left(\frac{|k|}{\lambda}\right)\right) \right] \end{aligned} \quad (4)$$

for all $k \in \mathbb{R}$. The parameters have the following meaning (see figure 4.1):

- μ is the location parameter.
- $c > 0$ is the scale parameter.
- $\alpha \in (0, 2] \setminus \{1\}$ is the characteristic exponent, which defines the overall shape of the PDF, and especially the fatness of the tails. To give an intuition, the closer α is to 2, the closer the TSL approximates a Gaussian.
- $\lambda > 0$ is the cutoff parameter, which is responsible for making the moments of the TSL well-defined.
- $\beta \in [-1, 1]$ is the skewness parameter: the PDF of the TSL is skewed to the right for negative β and to the left for positive β .

The PDF of the distribution can be obtained by fourier transforming the characteristic function:

$$\text{pdf}_{TSL}(x; \mu, c, \alpha, \lambda, \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-ikx - \psi(k; \mu, c, \alpha, \lambda, \beta)) dk \quad (5)$$

The fundamental property of this distribution which makes it well suited for the task of estimating

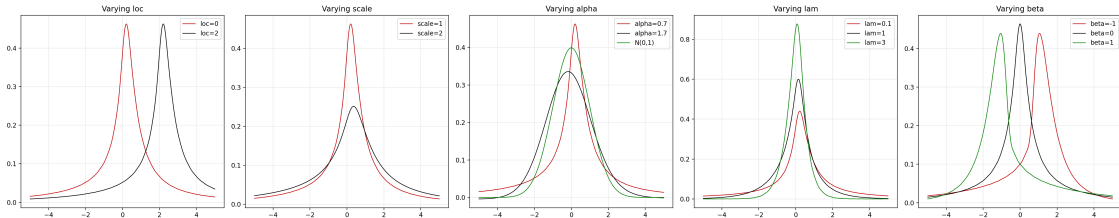


Figure 1: TSL PDF. Baseline parameters: $\mu = 0$, $c = 1$, $\alpha = 0.7$, $\lambda = 0.2$, $\beta = -0.2$

longer horizon variances is its convolution property. In particular the T -day convolution of the PDF satisfies:

$$\text{pdf}_{TSL}^T(x; \lambda) = T^{-1/\alpha} \text{pdf}_{TSL}\left(T^{-1/\alpha}x; T^{1/\alpha}\lambda\right). \quad (6)$$

Therefore, by fitting the TSL distribution parameters on daily returns and appropriately scaling the PDF, we can approximate the distribution of larger horizon returns.

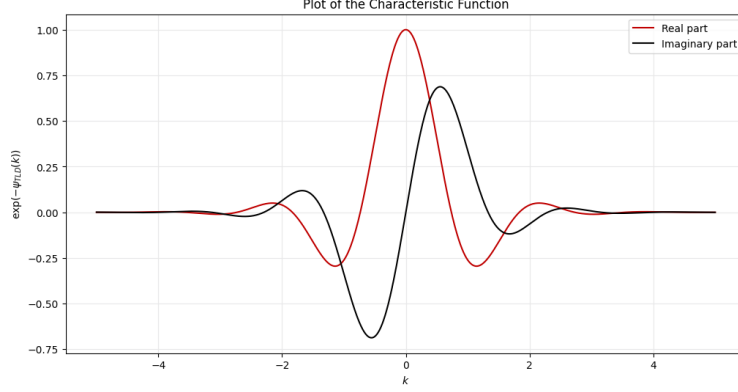


Figure 2: The characteristic function of the TSL. Parameters: $\mu = 2$, $c = 5$, $\alpha = 0.7$, $\lambda = 2$, $\beta = -0.8$

4.2 Fitting the TSL

To obtain the PDF, the characteristic function (see figure 2) must be numerically Fourier-transformed. Following the paper approach, we use Romberg integration for $k \in [-15, 15]$ using $2^{12} + 1$ steps. This calculation is computationally expensive: for a given set of the five parameters, evaluating the PDF for 1000 x points takes around 5 seconds. As the dataset contains around 7000 observations, the calculation of the loglikelihood for a given parameter set takes around 35 seconds, or in other words, even a simple grid search over the parameters α, λ, β (as we are fitting on standardized returns, we can assume $\mu = 0, c = 1$) with a $7 \times 7 \times 7$ grid would take around 3.5 hours to complete. Therefore, I resorted to a faster, but less accurate approach:

1. I compute the KDE PDF of the standardized returns with bandwidth parameter given by Scott's rule of thumb:

$$b = n^{-1/5}$$

2. I evaluate the KDE PDF in a symmetric, log-spaced grid (denser around 0) of 201 points for $x \in [-15, 15]$.
3. I perform a grid search to find the set of parameters α, λ, β that minimizes the total squared error between the KDE PDE and the TSL PDF calculated with the given set of parameters on the points in the log-spaced grid, effectively reducing the number of required calls to the TSL PDF to 201 for each parameter set. The grid was chosen by taking 13 equally spaced points in the following ranges:

$$\alpha \in [0.75, 0.81], \quad \lambda \in [1.60, 1.80], \quad \beta \in [-0.18, -0.12].$$

In figure 3, we see a plot of the resulting TSL PDF and in table 2 we see the results of the grid search.

Unfortunately this approach is not free of consequences in terms of accuracy, as the total mean squared error (TMSE) of the KDE PDF greatly dominates the total squared error of the grid

search:

$$\begin{aligned}\widehat{TMSE}_{KDE} &= \frac{1}{2\sqrt{\pi}} \sum_{i=1}^{201} \widehat{\text{pdf}}_{KDE}(x_i) = 1.88, \\ \widehat{TSE}_{GS} &= \sum_{i=1}^{201} \left(\text{pdf}_{TSL}(x_i, \hat{\alpha}, \hat{\lambda}, \hat{\beta}) - \widehat{\text{pdf}}_{KDE}(x_i) \right)^2 = 0.005,\end{aligned}\tag{7}$$

and the errors in table 2 are likely to severely underestimate the actual standard errors of $\hat{\alpha}, \hat{\lambda}, \hat{\beta}$ if treated as estimators of the parameters of the actual distribution of the data.

Parameters	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\beta}$
Values	0.765	1.67	-0.155
Absolute Errors	0.005	0.01	0.005
Standard Errors	0.002	0.006	0.002

Table 2: The SE have been calculated assuming a uniform distribution in the interval where the estimate lies, and thus dividing the absolute errors by $\sqrt{3}$. Note: These errors refer to the parameter values used to fit the KDE approximation, and do not reflect confidence intervals for the true underlying distribution.

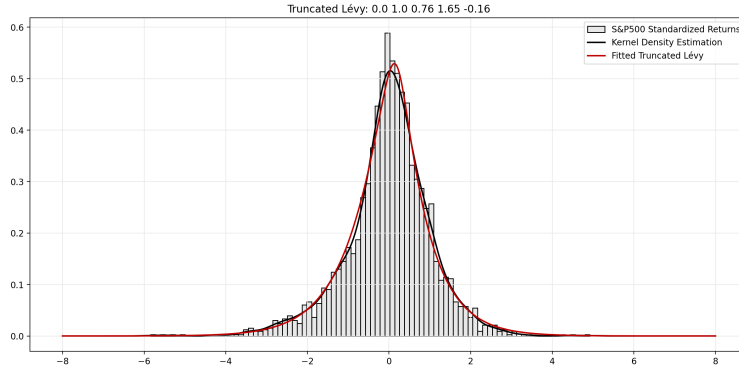


Figure 3: TSL PDF obtained by calibrating the parameters over the KDE PDF, and the standardized returns.

4.3 Estimating the 1-Day VaR

I then proceed to calculate the 1st, 5th and 10th percentiles of the (1-Day) distribution $D(0, 1) = TSL(0, 1, \hat{\alpha}, \hat{\lambda}, \hat{\beta})$ numerically:

$$Q(0.01) = -2.96, \quad Q(0.05) = -1.79, \quad Q(0.1) = -1.26$$

The $N\%$ 1-Day VaR is then obtained by multiplying the volatilities obtained from the GARCH by the N th percentile of the distribution.

$$VaR_t(N\%) = -Q(N\%) \sigma_t.\tag{8}$$

We define a *breach* a trading day when the realized return r_t is more negative than the minus VaR:

$$r_t < -VaR_t(N\%)\tag{9}$$

For the VaR estimate to be valid, over a time interval large enough the percentage of breach day should tend to $N\%$. In figure 4.3 we can see the minus 1% 1-Day VaR for the S&P500 compared to its daily returns. We achieved a reasonable amount of breaches in both the training (on which the model is calibrated) and the test set. The results are shown in table 3.

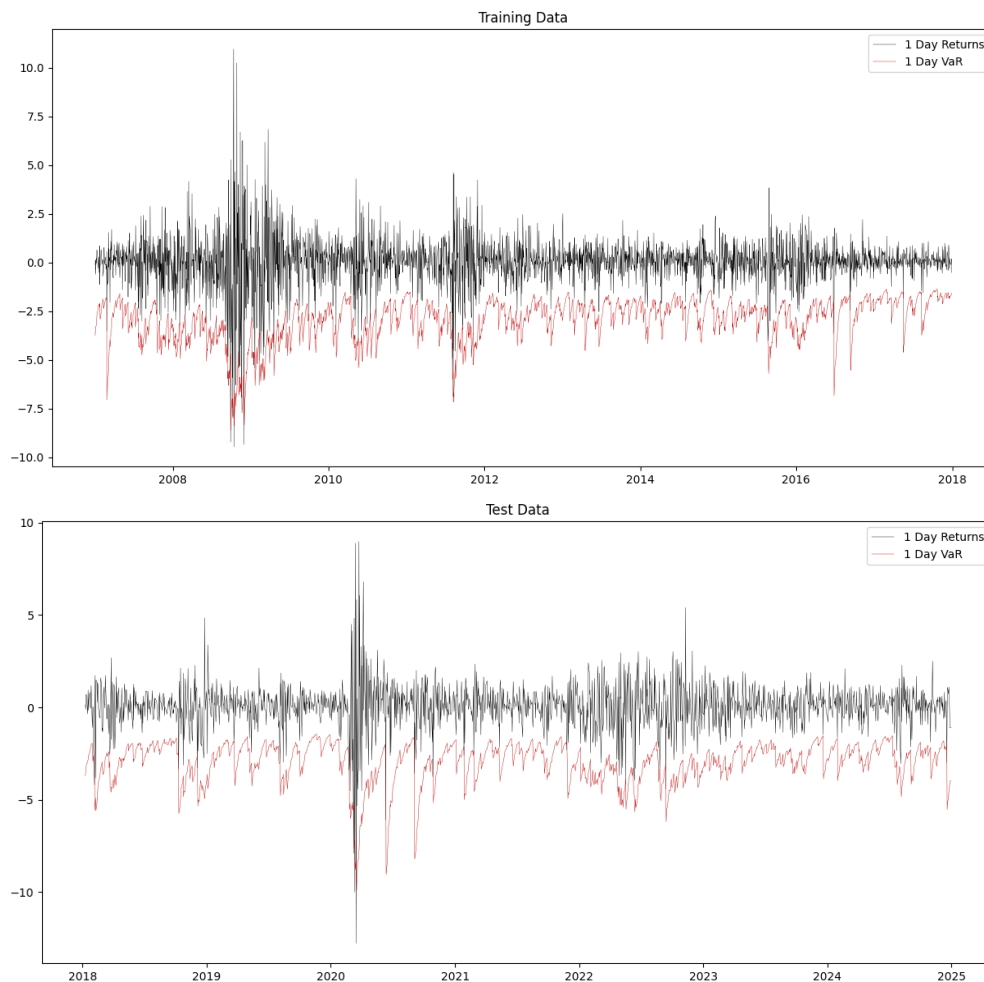


Figure 4: S&P500 1% 1-Day VaR and its daily returns

VaR	Training	Test
1%	0.94%	1.02%
5%	5.16%	4.56%
10%	9.82%	9.54%

Table 3: Daily VaR breaches in the S&P500

4.4 Estimating the 10-Day VaR

A similar procedure is repeated to estimate the 10-Day VaR. Taking advantage of the TSL convolution properties, we evaluate the quantiles of $D(0, N^{1/\hat{\alpha}}) = TSL(0, 1, \hat{\alpha}, N^{1/\hat{\alpha}}, \hat{\lambda}, \hat{\beta})$, to find:

$$Q_{10}(0.01) = -8.11, \quad Q_{10}(0.05) = -5.56, \quad Q_{10}(0.1) = -4.26$$

We thus calculate the 10-Days VaR by multiplying the quantile to the daily volatility, and we calculate the breaches, which are found in table 3.

VaR	Training	Test
1%	1.05%	2.22%
5%	4.01%	5.58%
10%	6.7 %	8.68%

Table 4: 10 Days VaR breaches in the S&P500

5 Direct Method

In this section I get rid of any assumption on the standardized return distribution, and opt for a full nonparametric approach, by directly estimating the empirical quantiles from the standardized data:

$$\hat{Q}_p(x) = x_{[h]} + (h - [h])(x_{[h]} - x_{[h]}), \quad h = (N - 1)p + 1. \quad (10)$$

This estimator is consistent and asymptotically normal (David and Nagaraja 2003):

$$\sqrt{N} \left(\hat{Q}_p(x) - x_p \right) \xrightarrow{d} N(0, \sigma^2), \quad \sigma^2 = \frac{p(1-p)}{f^2(x_p)} \quad (11)$$

where f is the PDF at the point x_p and can be estimated using KDE.

5.1 Estimating the 1-Day VaR

For daily data we already have the standardized returns, and the result in terms of breaches can be seen in table 5. With this method, associating a confidence interval to the results is straightforward,

VaR	Quantile	Training	Test
1%	-2.97	0.94%	1.02%
5%	-1.85	4.91%	4.22%
10%	-1.28	9.61%	9.46%

Table 5: Daily VaR breaches in the S&P500

as we can estimate quantile variance as described above, and volatility variances by Monte Carlo simulation of equation (1), with ϵ_t sampled from the KDE distribution.

5.2 Estimating the 10-Day VaR

To estimate the 10 days VaR, we start by recalibrating the GARCH model on the 10 days returns, standardize the returns based on the 10-days volatilities and then follow the same procedure as in the daily case 5.1 to obtain quantiles and hence the VaR.

VaR	Quantile	Training	Test
1%	-4.37	1.08%	1.13%
5%	-2.25	3.97%	3.98%
10%	-1.78	9.74 %	7.39%

Table 6: 10 Days VaR breaches in the S&P500

6 Conclusion

Both methods yield a notable improvement when compared to the naive model that assumes a normal distribution of returns. In terms of predicting the 10 Days VaR each has its advantages and disadvantages: The main Cons of the *scaling the TSL* approach are:

- Estimating the shape parameters of the TSL is very computationally intensive.
- Resorting to fitting the TSL on the KDE poisons the results with its error.
- The results are still notably off from the expected, especially for lower percentiles.

Its Pros:

- Works even when data is scarce
- Does not require to fit a GARCH on the 10 days returns

Cons of the *empirical quantile* method:

- Empirical quantile estimation may be flawed if data is scarce, yielding unexpected results.
- Requires to perform a GARCH calibration on aggregates, which may suffer when data is scarce.

Its Pros:

- Yield overall more realistic results.
- Does not require any assumption on the standardized return distribution.

References

- Christoffersen, Peter F (2003). *Elements of financial risk management*. Academic Press.
- David, Herbert A and H N Nagaraja (2003). *Order Statistics*. 3rd. John Wiley & Sons.
- Jorion, Philippe (2006). *Value at risk: The new benchmark for managing financial risk*. 3rd. McGraw-Hill.
- Lehnert, Thorsten and Christian C.P. Wolff (2004). “Scale-consistent Value-at-Risk”. In: *Finance Research Letters* 1.4. Limburg Institute of Financial Economics, Maastricht University and Centre for Economic Policy Research, pp. 205–219. ISSN: 1544-6123. DOI: 10.1016/j.frl.2004.03.002.
- Pagan, Adrian R and G William Schwert (1990). “Alternative models for conditional stock volatility”. In: *Journal of econometrics* 45.1-2, pp. 267–290.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall/CRC.