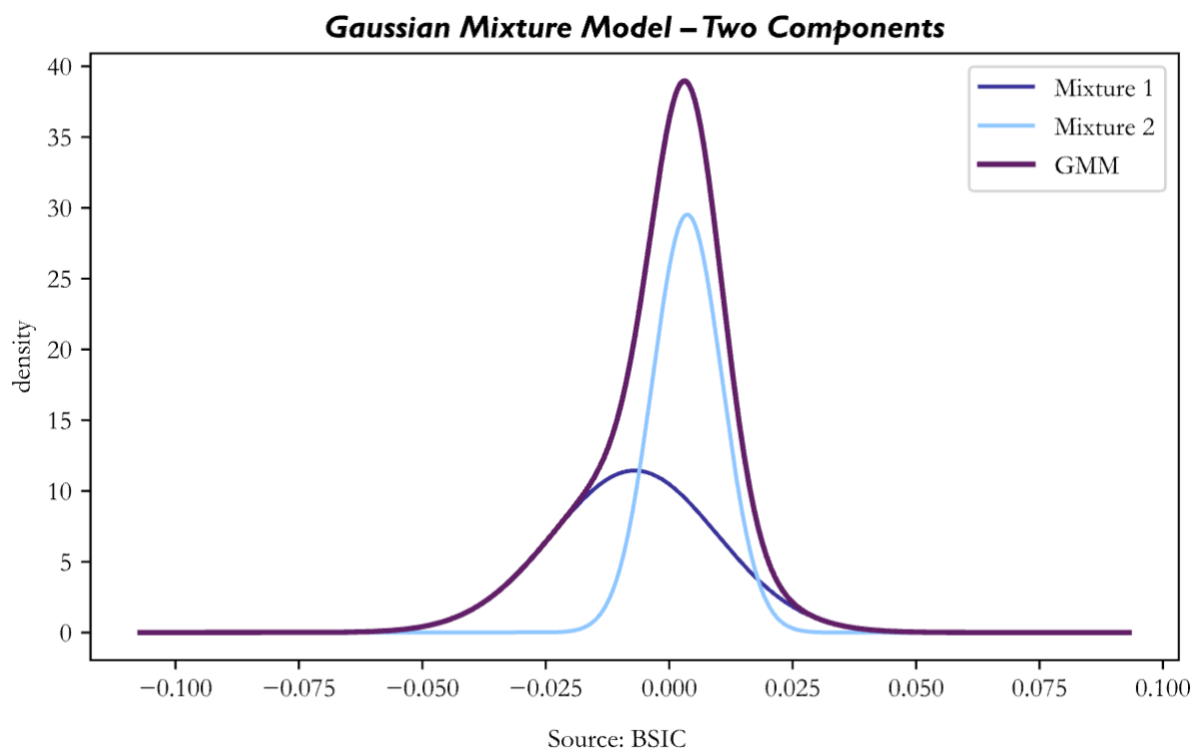


Regimes and Tails: A GMM Framework for VaR and ES

Introduction

Over the years many researchers studied returns in the market, in particular empirical analysis has shown that market returns follow a quasi-normal distribution. This assumption however doesn't capture the fact that markets exhibit more sharp movements than what the standard normal distribution would imply, thus we say that markets exhibit fatter tails.

To provide a better approximation of the actual market return distribution, we want to introduce the use of Gaussian Mixture Model (GMMs) for this scope. GMMs are models able to describe each continuous distribution with mixture components, which are multiple normal distributions having different weights. The weighted sum will then make up the final GMM distribution.



Each mixture component can be thought of as a market regime, with a latent variable that represents the current (active) regime. Across different market regimes, some studies [1] highlighted how the correlation among asset returns can change. A great example is during market downturns where the correlation among securities tends to rise. GMMs can thus help us to better approximate regime-dependent correlation structure.

In this article in particular we will dive deeper into the usage of Gaussian Mixture Models to better model the return distribution and to provide a more precise analysis of the risk that portfolios can exhibit, by utilizing the GMM distribution to compute both Value at Risk and Expected Shortfall.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

Entropy and GMM (mathematical intro)

Many observations suggest that entropy is preferable when dealing with complex systems since it better captures higher-order moments compared to standard deviation, given a more detailed outlook about the underlying volatility. In addition, entropy can measure non linearities and other complex dependencies that are not captured by common volatility measures.

To better understand the concept of entropy we first need to define what is a surprise. Imagine we have a pool full of red balls and a few blue balls, the probability of extracting a red ball will be higher than the one of extracting a blue ball. Therefore, we will be 'surprised' if the extraction results in a blue ball. Here we notice how surprise can be defined as:

$$\log\left(\frac{1}{p(x)}\right)$$

In this way the higher the probability the smaller will be the surprise. We can define entropy in a formal way as the measure of the average amount of information (or 'surprise') produced by a stochastic process, more directly entropy is a measure of the "average surprise" of a certain sum of events. By understanding this concept we can easily derive the common Shannon entropy formula given as:

$$H(X) = - \sum_{n=1}^k p(x_i) \log(p(x_i))$$

Moving to GMM, we can approximate the log-return with a finite mixture of Gaussian distributions defined as:

$$f(y; \theta) = \sum_{k=1}^G \pi_k \phi(y; \mu_k, \sigma_k)$$

where π_k refers to the probability of the G component, $\phi(y, \mu_k, \sigma_k)$ is the gaussian density with mean μ_k and standard deviation σ_k . All these components are store in the vector of random components $\theta = [\pi, \mu, \sigma]^T$.

It's important to notice that we use the standard deviation in our model by making the assumption of absence of autocorrelation, without this assumption covariance should be used.

To estimate this vector and fit the model we can either follow likelihood-based methods or Bayesian approaches. In this article we implement the EM algorithm, repeating the expectation and maximization steps until convergence is reached.

As for other distributions, we can compute the moments for the GMM that are given by:

$$\mu = E(Y) = \sum_{K=1}^G \pi_k \mu_k$$

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

$$\sigma^2 = V(Y) = E(Y^2) - E(Y)^2 = \sum_{k=1}^G \pi_k (\sigma_k^2 + \mu_k^2) - \mu^2$$

Higher moments can be derived using the same intuitions.

EM algorithm

Although the EM algorithm is not something easy to cover, it is useful to have an intuition on how hidden or unknown variables can be estimated using it. The process starts with an initial guess, then we have two steps that are Expectation and Maximization. In the expectation (E) we find missing data starting from the parameters and then we calculate the log-likelihood of the observed data using our current estimates. After this we proceed with the maximization step in which we maximize the log-likelihood and find new parameters. These two processes of changing parameters, calculate the log-likelihood and optimizing it, is repeated until we observe a converge or until the increase in the log-likelihood becomes negligible.

Entropy for GMM

This section will be quite math heavy. Estimating entropy for GMM can be quite challenging, however we will try to provide a direct way on how to do it. First, we can re-express the GMM introducing a multinomial discrete latent variable $Z \in 1, \dots, G$ and converting it to:

$$Z \sim M(1, \pi)$$

$$Y | (Z = K) \sim N(y, \mu_k, \sigma_k)$$

Although it seems difficult, the intuition is understandable, we are simply dividing the process in two steps: first we set us in the k distribution with the associated probability and then we see the Y conditioned to the fact that we are in that k distribution with their relative characteristics as mean and volatility.

Before deriving the final estimation of entropy recall that:

$$E(g(x)) = \int p(x)g(x) dx$$

Therefore:

$$E[\log(f(y)) | Z = K] = \int p(y|Z = K) \log(f(y)) dy = \int \phi(y, \mu_k, \sigma_k) \log f(y) dy$$

After this consideration we can derive the final formula given as:

$$\begin{aligned} H(Y) &= -\mathbb{E}_Z [\mathbb{E}_{Y|Z} [\log f(y)]] = -\sum_{k=1}^G \pi_k \mathbb{E}_Y [\log f(y) | Z = k] \\ &= -\sum_{k=1}^G \pi_k \int_Y \phi(y; \mu_k, \sigma_k) \log \phi(y; \mu_k, \sigma_k) dy. \end{aligned}$$

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

One disadvantage of entropy with respect to SD is that it is unitless, therefore it is difficult to interpret it. However, we can solve this by converting it into a standard deviation scale using the following expression:

$$\sigma = 2\pi^{-1/2} \exp(H(Y) - 1/2)$$

Model Selection

Since GMM distribution is made of a number of underlying mixture components we have to find the best number of them, finding ourselves in a trade-off between accuracy, which would asymptotically mean having one normal distribution to fit each single return with $\sigma = 0$, and model complexity, which would lead to overfitting instead of capturing the actual underlying distribution characteristics.

In order to find the most efficient number of components we rely on an indicator commonly used to identify the best model as the one that minimizes the Bayesian Information Criterion. It rewards models that explain data well, but penalizes the usage of too many parameters, thus by minimizing it we can find the most optimal equilibrium between the two.

The BIC is defined as:

$$\text{BIC}_M = -2\ell_M(\hat{\theta}) + v_M \log(n)$$

Where:

- $-2\ell_M(\hat{\theta})$ represents the goodness of fit of the model, defined as the maximized log-likelihood.
- $+v_M \log(n)$ represents instead a penalty component for complexity, with v_M denoting the number of independent parameters to be estimated, while $\log(n)$ is simply the logarithm of the sample size.

Thus, during our model selection process we will select the model showcasing the lowest BIC value that is the one maximizing the payoff described above.

GMM vs Gaussian Normal Distribution

As discussed above, the Gaussian Mixture Model (GMM) provides a more flexible representation of asset-return distributions than a single-Gaussian benchmark. To validate this, we fit both the optimal GMM and a single Gaussian to the NASDAQ Composite, Bitcoin, Crude Oil, and Euro Stoxx 50, using a five-year in-sample window. Model performance is then evaluated on a one-year out-of-sample period, under the standard assumption that daily log returns exhibit negligible serial autocorrelation.

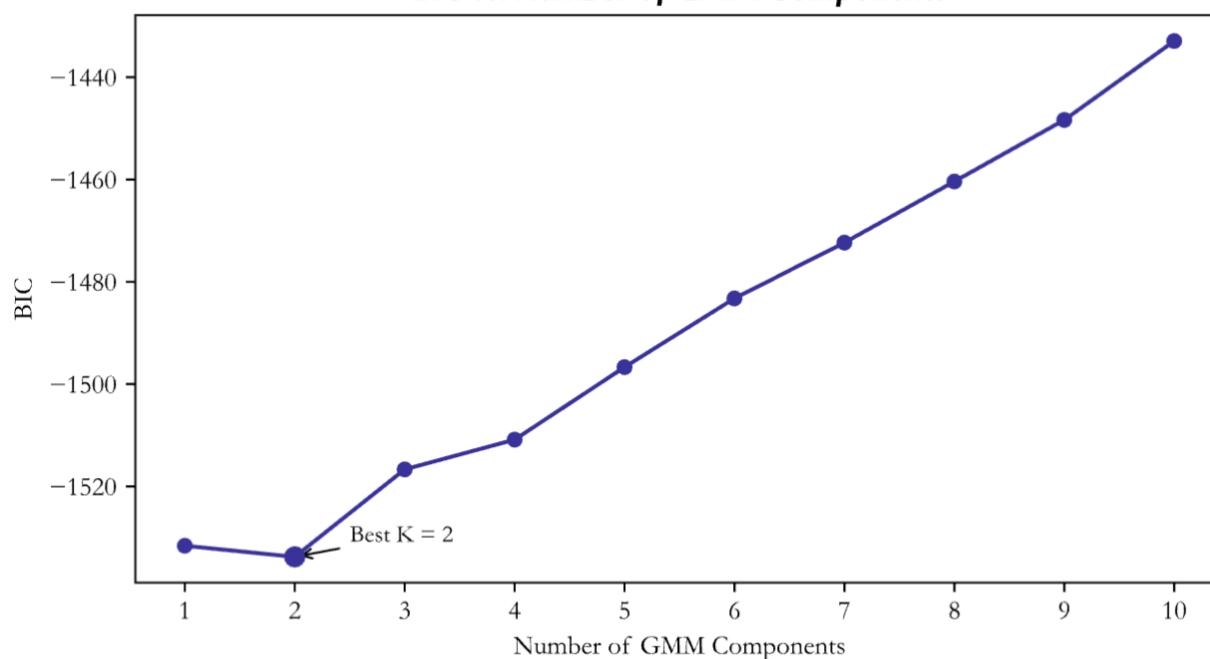
To compare models, we compute each model's out-of-sample log-likelihood and define their difference, Δ (GMM minus Gaussian). By construction, $\Delta > 0$ indicates that the GMM provides the superior fit, whereas $\Delta < 0$ favors the Gaussian benchmark. We estimate GMMs with one to ten components and, for illustration, report the corresponding Bayesian Information Criterion (BIC) for the NASDAQ Composite, selecting the specification that

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

minimizes

the

BIC.

BIC vs. Number of GMM Components

Source: BSIC

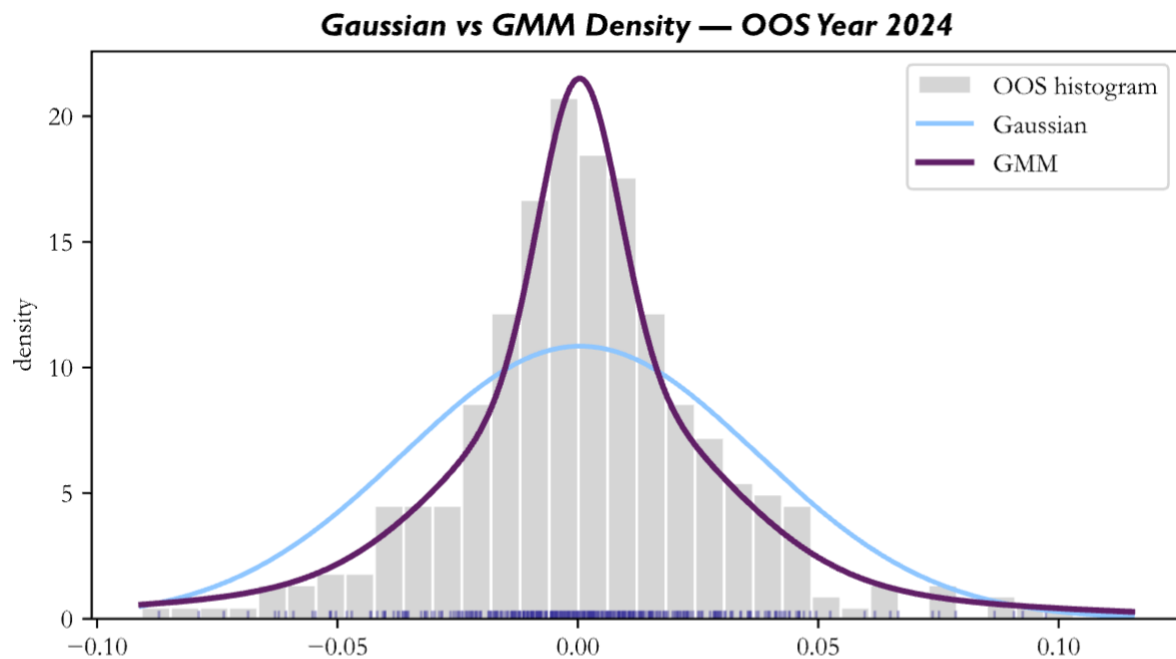
The Bayesian Information Criterion (BIC) reaches its minimum at two mixture components and generally maintains this characteristic in all the backtests, confirming that the return distribution is best represented by a bimodal structure. Intuitively, each component can be interpreted as a distinct market state: one corresponding to periods of low volatility and stable returns, and the other to high-volatility episodes typically associated with market stress or sharp price adjustments.

This interpretation intuitively aligns with how markets work, tending to oscillate between calm and turbulent phases, and the GMM framework captures these shifts endogenously, without the need to impose explicit regime-switching assumptions.

The results confirm that the Gaussian Mixture Model consistently provides a superior fit to market return distributions across all assets considered. Among them, Bitcoin—characterized by pronounced volatility and heavy tails—shows the most substantial improvement when modelled through a GMM rather than a single Gaussian distribution.

As illustrated in the comparative density plot below, the two-component GMM trained on return data of Bitcoin from 2019 to 2023 aligns closely with the empirical return distribution, effectively capturing both the central mass and the tail behaviour that the normal model systematically underestimates. This visual evidence reinforces the statistical findings, underscoring the GMM's ability to represent complex market dynamics more accurately than traditional Gaussian assumptions.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

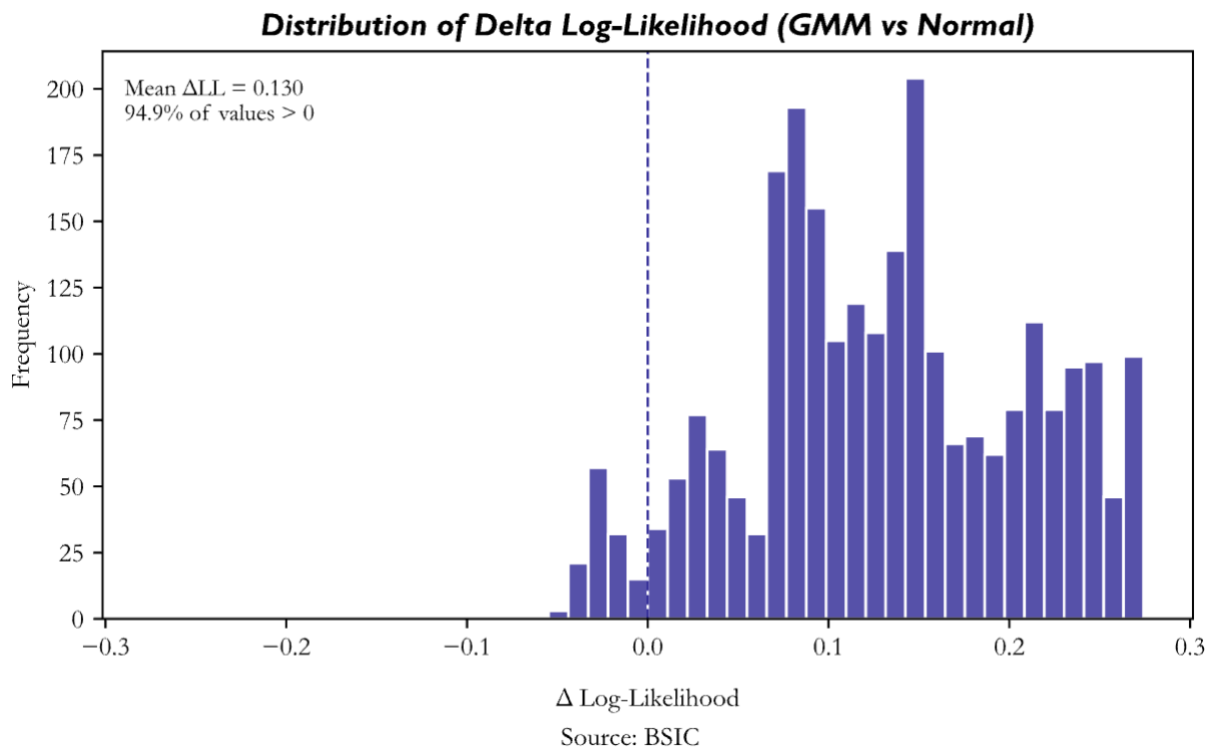
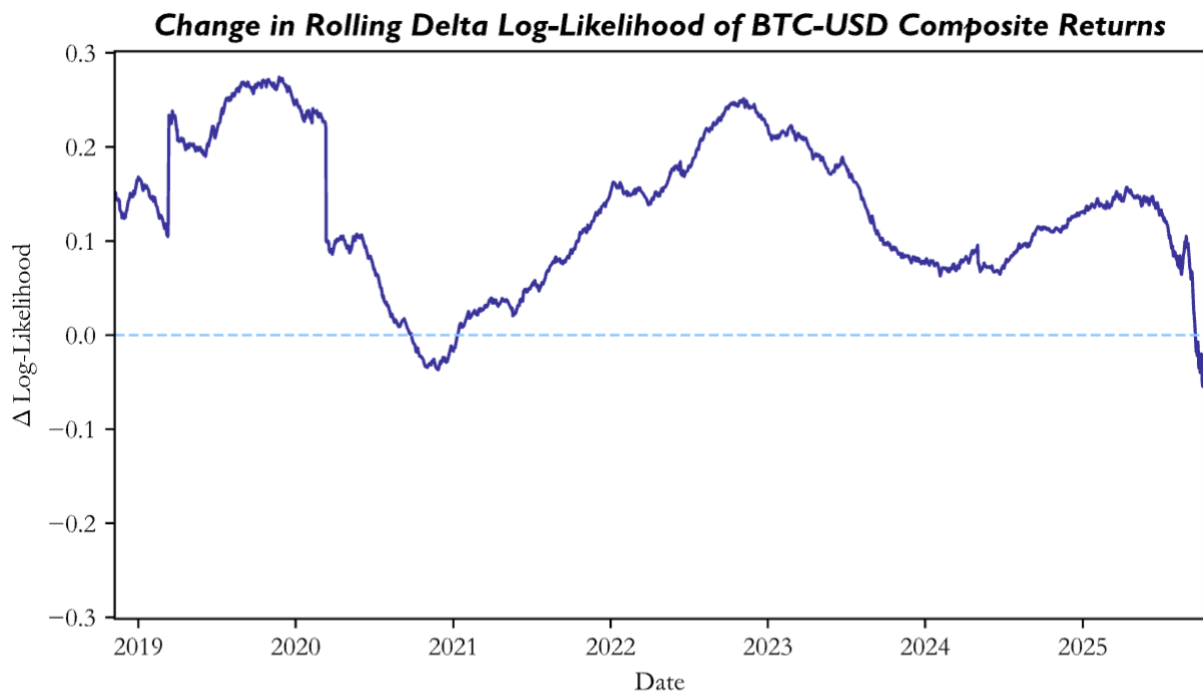


Other than graphically we empirically analysed the day by day Δ Log Likelihood on Bitcoin returns, which exhibits a positive mean of 0.13 with 94.9% of the observation exhibiting a positive Δ . This once again confirms our take that the GMMs can be applied to fit very well the actual returns.

As shown in the upper chart, the Δ log-likelihood remains predominantly positive throughout the sample, indicating that the GMM consistently achieves a higher likelihood of observing the realized returns.

The lower histogram complements this view by displaying the distribution of Δ values, with a mean of 0.13 and 94.9 % of observations above zero. Together, these results confirm the persistent superiority of the GMM specification in modeling Bitcoin's return dynamics, reinforcing the evidence previously observed in the density comparison.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

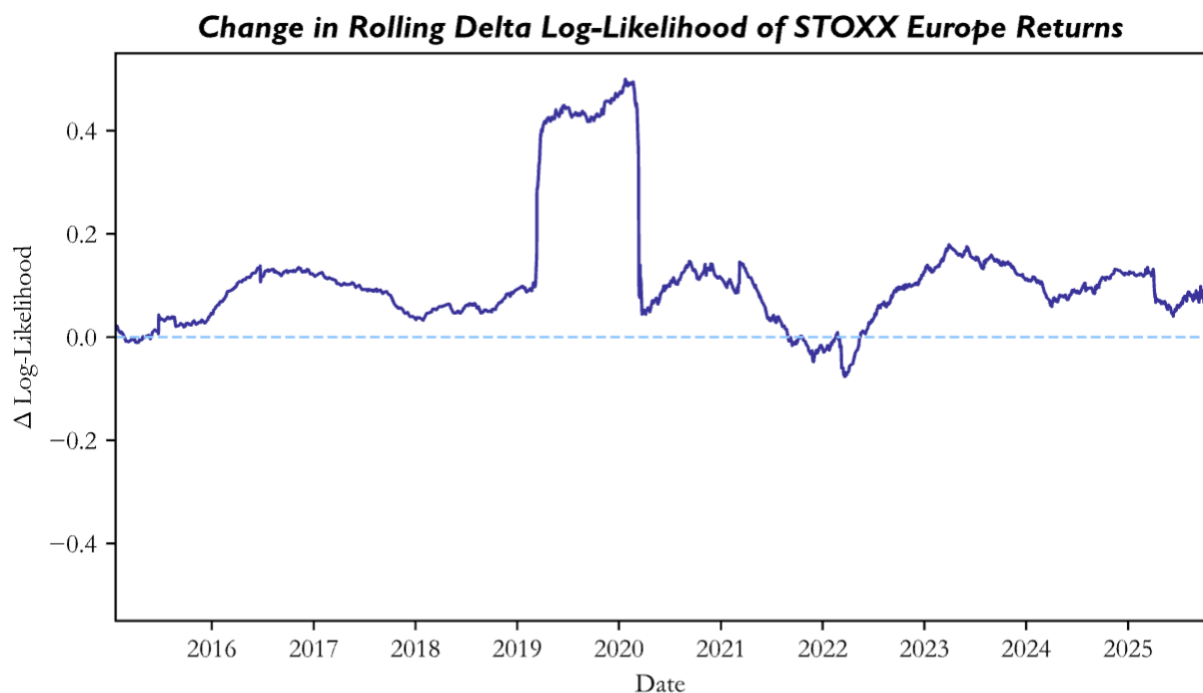


All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

To validate the robustness of these findings, the same procedure was extended to a broader set of assets, including the NASDAQ Composite, Gold, Bitcoin, and EuroStoxx 50. Across all cases, the GMM outperformed the standard Gaussian model, consistently delivering higher log-likelihood values in out-of-sample tests.

The Bayesian Information Criterion (BIC) generally favoured models with two mixture components, suggesting the bimodal market regimes even in more traditional asset classes. While the magnitude of improvement varied across instruments, the direction of the results was uniform, the GMM offered a more accurate and flexible representation of return distributions, particularly in capturing fat tails and asymmetries often overlooked by single-Gaussian specifications.

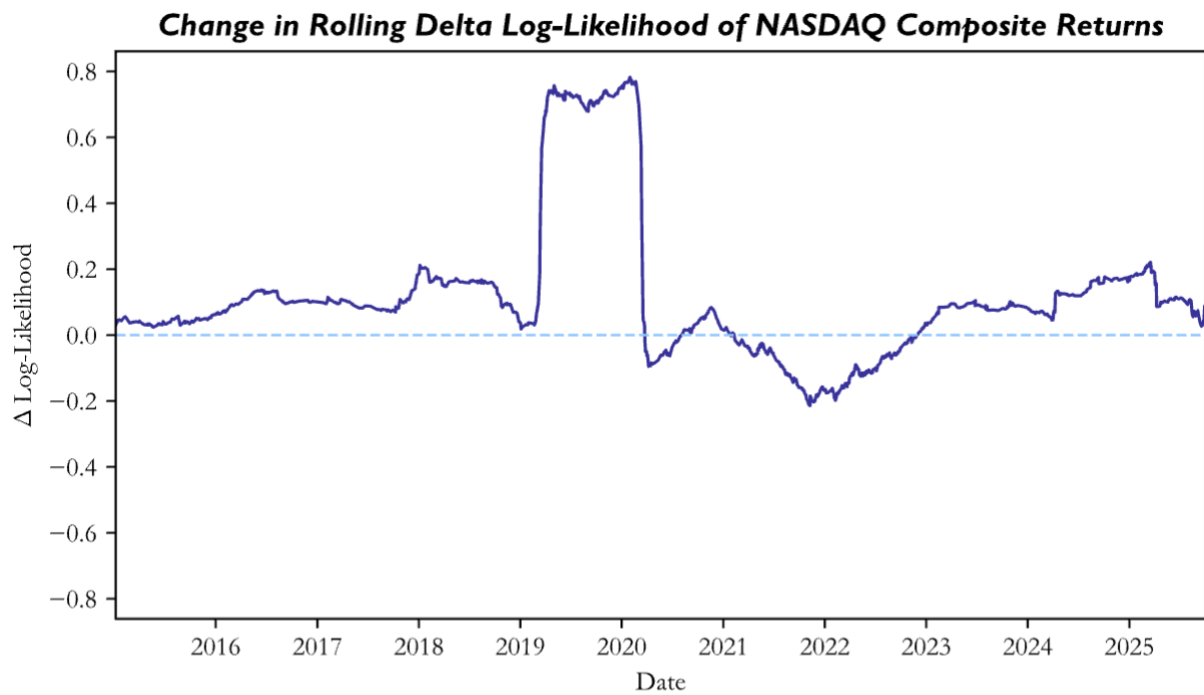
The rolling Δ log-likelihood for the EuroStoxx 50 shows that the GMM generally outperforms the Gaussian model, with a mean Δ log-likelihood of 0.113 and approximately 92% of observations above zero.



Source: BSIC

The rolling Δ log-likelihood for the NASDAQ Composite shows that the GMM generally outperforms the Gaussian model, with a mean Δ log-likelihood of 0.116 and approximately 80% of observations above zero.

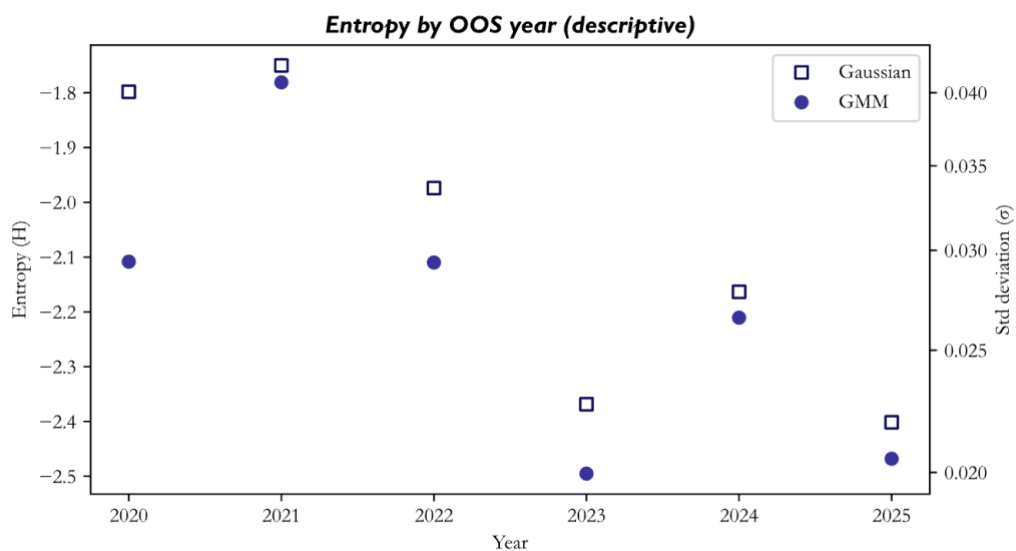
All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.



Source: BSIC

Results on Entropy

We estimate yearly differential entropy of out-of-sample returns under two models: a single Gaussian and the GMM selected by BIC. Entropy measures the average surprise of the return-generating process; for a Gaussian it is monotone in σ (higher $\sigma \Rightarrow$ higher H). For general, non-Gaussian returns we adopt the mixture-based estimator, evaluating the fitted GMM density at observed returns, which avoids binning/bandwidth choices and directly reflects the mixture structure.



Source: BSIC

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

From these analytical returns we see how GMM models exhibit more negative values, than the Normal Gaussian counterpart, this happens because while the GMM distribution can better adapt to central data clustering and fat tails, the Normal distribution cannot and thus to better fit fat tails it will overestimate standard deviation thus providing a worse prediction of the central cluster of returns.

A mixture, instead, concentrates probability mass around the relevant regime modes, so the average “surprise” of realized returns is lower. Analytically, for two-component mixtures, as component means separate, variance can keep rising while entropy plateaus—the model becomes more dispersed but not more uncertain. Hence, GMM entropies tend to be lower (more negative) than Gaussian entropies when regimes are present.

VaR and Expected Shortfall

Value at Risk (VaR) and Expected Shortfall are too widely used measures for assessing the potential risk of loss in a portfolio. Both are very useful since they give us a fast and easy way to interpret the magnitude of a loss that could happen. Starting from the VaR we can define it as:

$$VaR(\alpha) = \inf l \in R: P(L > l) \leq \alpha$$

This can be seen as the lower bound of the set of losses such that the probability of having a random loss greater than our loss is less than alpha. Intuitively we set a threshold by which we are confident that only with little probability we will suffer a loss greater than it. For example, a \$5,000 one day VaR at a 99% confidence level for a portfolio of \$100,000 means that there is a 1% chance that we will lose more than \$5,000 in one day.

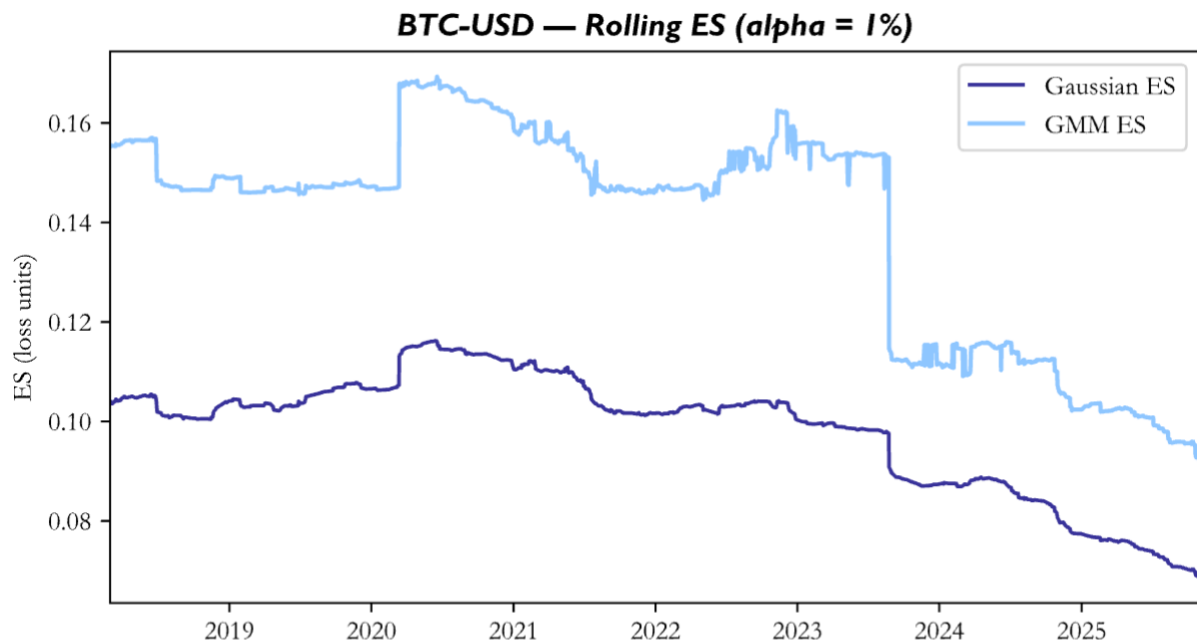
On the other hand, the ES is the expected loss given that the loss exceeds VaR. Meaning how much we expect to lose if things go bad. The ES formula can be written as:

$$ES(\alpha) = E[L|L \geq VaR(\alpha)] = \frac{1}{\alpha} \int_0^\alpha VaR(u)u$$

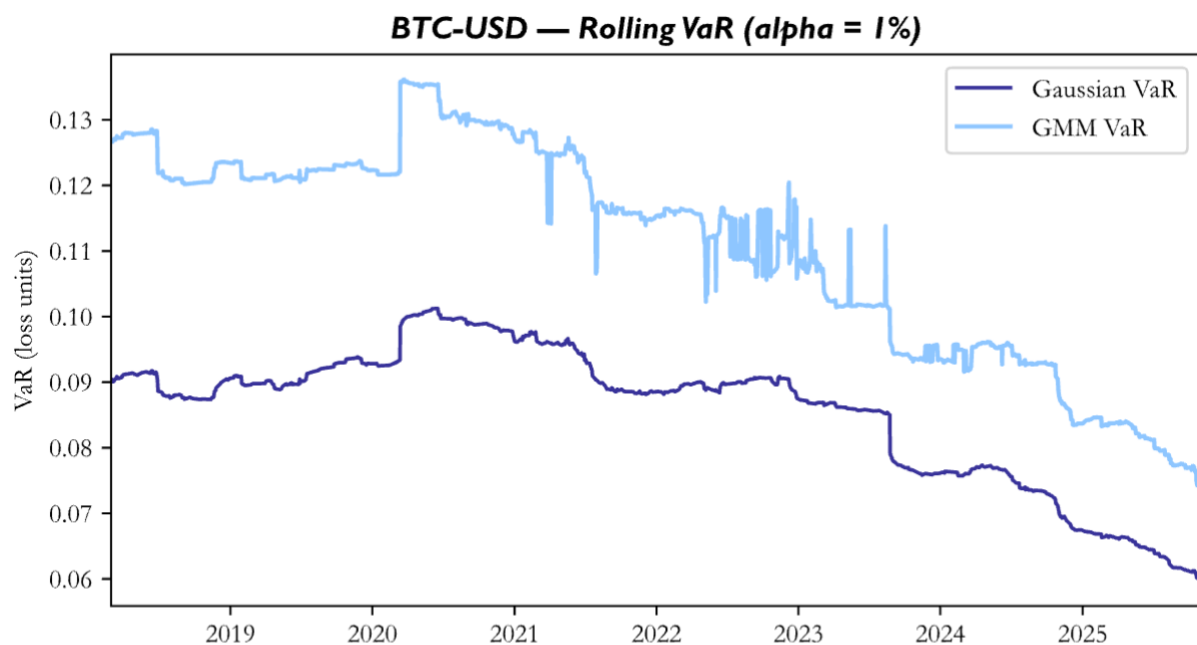
Building on the density results, we translate model differences into tail-risk measures by computing 1% Value at Risk (VaR) and 1% Expected Shortfall (ES) from each model’s one-day-ahead predictive density. As in the earlier sections, parameters are estimated on a five-year rolling window and the forecasts are evaluated out of sample day by day, under the standard assumption that daily log-returns display negligible autocorrelation. VaR is the upper α -quantile of the loss distribution, while ES is the conditional mean loss beyond that quantile; for the GMM these are obtained from the fitted mixture cdf and its implied tail expectation, avoiding simulation and respecting the mixture’s heavy-tail structure.

The first two plots report rolling ES and rolling VaR for BTC-USD at $\alpha = 1\%$. The mixture-based series move in discrete steps and re-level rapidly when the return distribution switches regime, whereas the single-Gaussian benchmarks adjust more gradually. For the whole sample, and especially around turbulent episodes, the GMM lines sit above the Gaussian lines, signalling fatter estimated left tails and therefore more conservative tail-risk forecasts when markets are stressed. As conditions normalize, the two specifications converge, which is coherent with a single-regime environment where a Gaussian approximation is adequate. This visual pattern is precisely what one would expect given the mixture’s ability to capture asymmetry and tail thickness that a single Gaussian cannot reproduce.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.



Source: BSIC

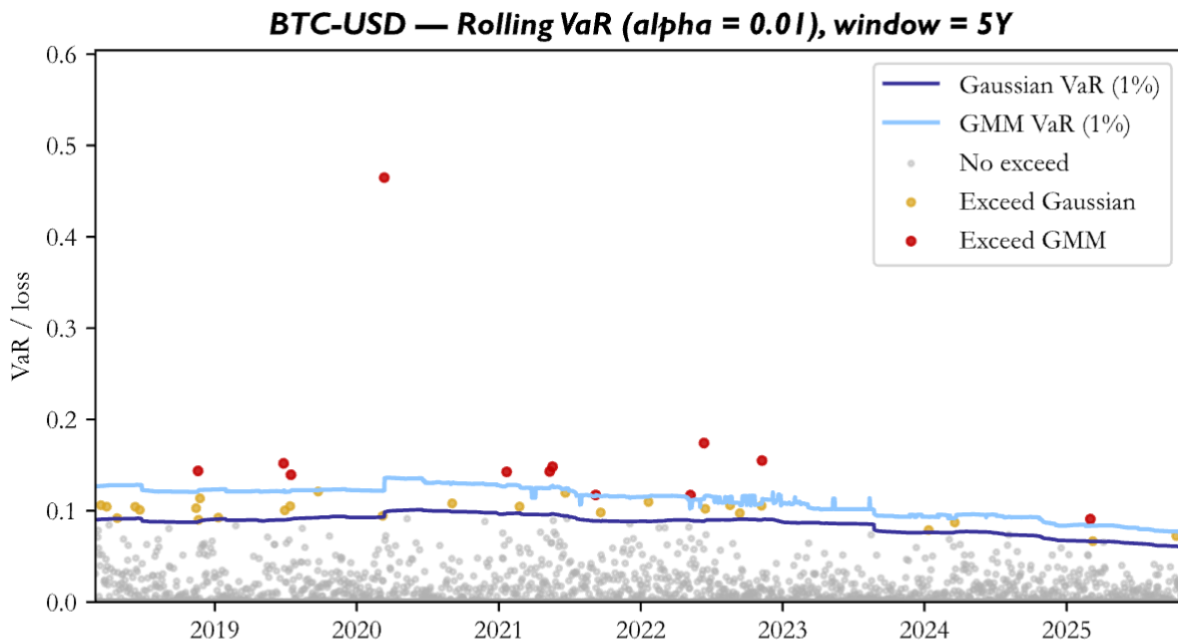


Source: BSIC

This graph overlays realised daily losses on the rolling 1% VaR forecasts and marks exceedances. Points above the Gaussian VaR appear more frequently than those above the GMM VaR. In particular, Gaussian quantile tends to be crossed more often during stress, while the GMM quantile is crossed less frequently and with less bunching.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

This is consistent with the mixture assigning more mass to the far-left tail when regime shifts occur, thereby reducing unexpected breaks of the risk threshold. It also aligns with the earlier Δ log-likelihood analysis, which indicated that the mixture offers a superior out-of-sample description of the returns across time; the tail-risk plots show that this distributional advantage propagates to risk measures, not only to central-density fit.



Source: BSIC

Similar results can be obtained in other assets.

Conclusion

We introduced a Gaussian Mixture framework as a flexible alternative to the single-Gaussian benchmark, selected the number of components via BIC, and estimated models on rolling five-year windows with one-day-ahead evaluation. Linking density to risk, we computed 1% VaR and 1% ES directly from each model's predictive distribution and used entropy, mapped to a volatility-equivalent scale, to interpret uncertainty.

Empirically, the GMM provides a superior out-of-sample fit (positive Δ log-likelihood), lower entropy in regime-rich years, and more responsive VaR/ES during stress, converging toward the Gaussian in calm periods. Visual VaR exceedance overlays indicate fewer, less clustered breaches under the mixture.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

References

- [1] Ang, A.; Bekaert, G., “How Do Regimes Affect Asset Allocation?”, 2004
- [2] Scrucca, Luca, “Entropy-Based Volatility Analysis of Financial Log>Returns Using Gaussian Mixture Models.” 2024

TAGS: GMM, Entropy, Tail risk, Value at Risk (VaR), Expected Shortfall (ES), Regime switching, EM algorithm

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgment at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.