Journal Pre-proof

Volatility forecasts by clustering: Applications for VaR estimation

Zijin Wang, Peimin Chen, Peng Liu, Chunchi Wu

PII: S1059-0560(24)00332-0
DOI: https://doi.org/10.1016/j.iref.2024.05.034
Reference: REVECO 3355

To appear in: *International Review of Economics and Finance*

Received date : 7 March 2024
Revised date : 8 May 2024
Accepted date : 9 May 2024

Please cite this article as: Z. Wang, P. Chen, P. Liu et al., Volatility forecasts by clustering: Applications for VaR estimation. *International Review of Economics and Finance* (2024), doi: https://doi.org/10.1016/j.iref.2024.05.034.

Volatility Forecasts by Clustering: Applications for VaR Estimation

Abstract

It is well known that volatility has time-varying and clustering characteristics. The information content of volatility clustering is particularly important in turbulent periods, such as the stage of financial crisis. How to fully mine the implicit information within clusters to predict the volatility in the future is a rarely discussed issue. In this paper, we put forward a partition model to segment volatility into non-overlapping clusters by Fisher's optimal dissection methodology. Using this model, we can quickly identify the points of structural changes in volatility. By utilizing the information of the nearest cluster, we can perform point estimation and interval estimation on future volatility. In the end, we conduct some empirical examples based on the returns of S&P 500, DAX 30 and FTSE 100 index. We find that our method can improve the volatility forecast and VaR estimations.

Keywords: Volatility forecasts; Fisher's optimal dissection; value-at-risk.

1. Introduction

Return volatility is crucial in asset pricing, portfolio selection and risk management (Angelini et al., 2019; Schmitt and Frank, 2017; Engle and Siriwardane, 2018; Chen et al., 2019). The classic Autoregressive Conditional Heteroskedasticity (ARCH) model and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model proposed by Engle (1982) and Bollerslev (1986) are popular for their simplicity and easiness of interpretation in volatility. Despite their extensive applications, GARCH-type models still struggle with rapid market structural changes (Andersen et al., 2003; Heston, 2012; Smetanina and Wu, 2021), which leads to forecasting challenges. For instance, when a new small sample emerges, indicating a sharp drop or surge, the model assimilates it within the entire data set. Given this, a single new small sample can not substantially alter the estimated parameters. Consequently, the model is hard to determine the nature of this change whether it's anomalous outliers or a genuine shift in volatility. Even when

considering continuous-time dynamic volatility models, such as Heston and CEV models, we still have a lag problem in parameter estimation. To address these challenges, this paper introduces a volatility cluster partition model using clustering analysis to forecast volatility and improve the Value-at-Risk (VaR) metric as an application. The results demonstrates significant improvements in performance over traditional methods.

In stock market, stock prices may suddenly change usually due to external factors or facing catastrophic risks. Accurate volatility prediction is crucial for stock or option trading strategy especially in financial crisis. For instance, the circuit breaker mechanism was rarely triggered until 2020. On October 27, 1997, the Dow Jones Industrial Average index plummeted 7.18%, its largest decline since 1915 exceeding the upper limit of 7%. From March 9 to March 18, 2020, the circuit breaker was triggered four times. Particularly on March 16, 2020 the S&P 500 index's over 7% drop led to a trading halt, the Dow Jones suffered its most significant daily decline in 33 years and dived by 11.98%. The characteristics of this period are intensified market volatility and structural changes. Overlooking these changes in time series could result in inefficient volatility forecasts. Rapidly identifying and incorporating these structural shifts in volatility is crucial for understanding market risks and trading dynamics. Effective future volatility prediction requires a method that is both simple and also responsive to sudden market changes.

Schmitt and Frank (2017) posited that the speculators' herding behavior leads to volatility clustering, where volatility can remain stable for a period before abruptly switching status. During stable periods, dynamic variation is less crucial, and the mean of volatility becomes the key point. Thus the average value of volatility during these periods can effectively represent the overall volatility. A critical aspect is to identify the optimal times for clustering. In stable market conditions, investors tend to make independent decisions. However, in turbulent times, market participants are often swayed by others' behaviors, gravitating towards either buying or selling. This has led to an imbalanced market scenario, with significant adjustments in stock prices and increased volatility.

The market regime, whether calm or turbulent, generally exhibits persistence. Nonetheless, significant events, such as financial crises, wars, or major policy alteration, can prompt changes in these regimes. Regime-switching (RS) processes are used to model

these changes. Hamilton (1989) introduces a model depicting the U.S. business cycle's swifts between recessions and expansions. Klaassen (2002) proposes the RS-GARCH models, incorporating GARCH in each regime. Alternatively, Pelletier (2006) developed a model for multiple time series with regime-switching dynamic correlations, where covariance remains constant within a regime but varies across regimes. Signals of regimes switchings in RS models are often identified as structural change points (Andreou and Ghysels, 2002; Lee et al., 2015).

RS models, while useful, come with notable drawbacks, such as the complexity and high computational cost of parameter estimation. Additionally, accurately estimating the probability of status transition in these models is a challenge. In a fast-paced real market, the ability to quickly adapt to fluctuations is crucial. Consequently, there is a demand for a model that is not only easily estimable but also capable of integrating timely market information. Furthermore, the subjective nature of interpreting market changes, due to the transition probabilities, poses a challenge for investors. Therefore, there is an urgent need for a fast, timely, simple and objective method to describe volatility. Dealing with these challenges is our main motivation to develop our models.

To overcome the above difficulties, we can consider the volatility obtained by GARCH family models as an object that can be clustered, and use unsupervised learning clustering algorithms to classify the volatility obtained. In the past two decades, clustering analysis has become increasingly popular, with methods like k-means, mean-shift clustering (Cheng, 1995), and DBSCAN (Ester et al., 1996), which have extensive uses in fields from statistics (Jain, 1999) to image segmentation (Coleman, 1979). Advanced techniques involving parameters, such as support-vector machines and neural networks, have been introduced in artificial intelligence (Zhang, 2000; Shi et al., 2016), effectively handling non-sequential data. However, in financial modeling most of data is about time series, an alternative approach to clustering the time series data is necessary. Fisher's optimal dissection method is particularly suited for this point, enabling dynamic volatility clustering by minimizing a loss function based on sample similarity. This principle way has plentiful applications in economics (Bai and Perron, 2003) and image processing (Verbesselt et al., 2010) highlighting its versatility.

By employing Fisher's optimal dissection method, we can pinpoint the time points for volatility clustering and use representative volatility values from stable periods. When

analyzing a new data-set, it is easy to assess whether it is suitable for the latest cluster. If so, use the representative value of the cluster for future volatility prediction. If not, this may indicate the emergence of a new cluster.

Our model distinguish itself from conventional models in detecting structural changes, excelling at creating optimal volatility partitions and promptly classifying them using up-to-date data. In contrast to traditional methods that require substantial data for detecting structural changes, our model is adept at quickly assimilating market information with fewer observations. This efficiency stems from the fact that conventional tests largely depend on statistical inferences drawn from model parameters, revealing structural changes only after accumulating a significant number of samples. This limits their capacity to promptly identify cluster swifts in volatility. Our approach, a semi-parametric model, does not hinge on model parameters, allowing for direct assessment of structural changes in volatility.

GARCH models are widely used for volatility forecasting in academia and industry. Orhan and Köksal (2012) explored various GARCH models for VaR estimation during high-stress periods. VaR quantifies the maximum projected loss of an investment portfolio over a specific period at a set confidence level. Common methods for VaR estimation include nonparametric historical simulation and parametric variance-covariance models. In this study, we use GARCH as a benchmark to evaluate the performance of our proposed model in VaR application, and also consider other models like GJR-GARCH, HAR, and asymmetric HAR for a comprehensive comparison.

This paper provides several key contributions. First, we introduce a novel model for clustering dynamic volatility using Fisher's optimal dissection method. This model uniquely maintains consistent volatility within each cluster while quickly adapting to changes at cluster transitions. Second, our model reduces estimation costs significantly as it does not rely on transition probability, unlike other models. Third, it consistently outperforms traditional GARCH family models in volatility forecasting. Utilizing our forecast volatility, we demonstrate marked improvements in risk management, especially in VaR estimation. Furthermore, our model's cluster volatility more closely reflects actual market volatility compared with standard GARCH models.

Our model is not only used for estimating VaR but also it can be used in credit analysis. In the KMV model, the distance to default depends on the volatility of the

company's assets. Therefore, the volatility that is more suitable for the latest market will be the key to improve the estimation of distance to default. Therefore, our volatility partition method provides a useful perspective for improving the calculation of the estimated default rates. In portfolio theory, Sharpe ratio also depends on the volatility of the asset portfolio. Therefore, our volatility partition model also has very broad application prospects in the field of asset portfolios. We can adopt the new model in this article to construct a volatility cluster that is more suitable for the market, thereby improving the solution of classical portfolio optimization problems and making it more suitable for market changes. In addition, according to the Bessel Agreement, banks are required to retain a certain proportion of retention funds when making investments. This ratio is calculated based on VaR. Therefore, the most important application of VaR estimation in this article will provide a more market-oriented and powerful tool for financial regulation.

This paper is structured as follows. Section 2 introduces a range of dynamic volatility models as a basis of comparison. In Section 3, we present the cluster model based on Fisher's optimal dissection method. Section 4 explores the application of our volatility measures to VaR and discusses the enhancement of forecasts in dynamic volatility models. Section 5 presents details of our empirical results concerning volatility forecasting and VaR. We make a conclusion in Section 6.

2. Volatility clustering model

In this part, we provide a concise review of the classical GARCH family models. These models provide a basis for our model construction and comparison.

2.1. GARCH model

Consider a time series of returns r_t, $t = 1, \ldots, T$ follows,

$$\begin{aligned} r_t &= \mu_t + e_t, \\ e_t &= \sigma_t \varepsilon_t, \\ \varepsilon_t &\sim D(0,1), \end{aligned} \tag{1}$$

where μ_t denotes the dynamic mean and e_t represents the residuals, $\{\varepsilon_t\}$ is an independent and identically distributed innovation process with zero mean and unit variance.

$\sigma_t > 0$ is the conditional volatility independent of $\{\varepsilon_t\}$. The conditional variance, given by σ_t^2, is based on the past information $\mathcal{F}_{t-1}$ and is defined as $\sigma_t^2 = \mathbb{E}[r_t^2|\mathcal{F}_{t-1}]$.

For volatility clustering, it is evident that σ_t is not independent but correlated with both time and previous states. The initial models addressing volatility clustering are Autoregressive Conditional Heterogeneous (ARCH) model and its extension, the General Autoregressive Conditional Heterogeneous (GARCH) model (Engle, 1982; Bollerslev, 1986). A canonical GARCH(p,q) is expressed as

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

where ω is a constant parameter, α_i are the ARCH parameters and β_j are the GARCH parameters. The conditional variance is commonly parameterized with $p = 1$ and $q = 1$, which simplifies the equation to

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{3}$$

Later, various extensions of GARCH model have been proposed. One notable extension is the Glosten, Jagannathan and Runkle (GJR) GARCH model (Glosten et al., 1993). The GJR-GARCH model incorporates an additional term to account for the asymmetric response of volatility to positive and negative shocks, thus capturing the leverage effect more effectively. The model is shown as follows.

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 \mathbb{1}_{r_{t-1}<0} + \beta \sigma_{t-1}^2, \tag{4}$$

where ω, α, β are constant parameters as before, $\mathbb{1}_{r_{t-1}<0}$ is an indicator function and γ is the leverage parameter, which quantifies the different impact of negative shocks compared to positive ones. Positive γ indicates that negative returns lead to higher volatility more than positive returns.

2.2. HAR-RV model

Realized variance (RV) stands as a robust empirical metric, effectively gauging daily volatility of returns by leaning on cumulative squared intra-day returns. Mathematically,

it can be denoted as:

$$RV_t^M = \sum_{i=1}^{M} r_{t,i}^2,$$

where $r_{t,i}$ is the i-th logarithmic return during day t and M encapsulates the total number of returns in each day. As the granularity of data intensifies, that is M approaches infinity, RV evolves to be a consistent estimator of the true variance (Andersen and Bollerslev, 1998; Barndorff-Nielson and Shephard , 2002; Andreou and Ghysels, 2009).

In Hansen and Lunde (2006), a bias correction for realized volatility is presented. The adjusted formula is given by

$$RV_t^M = \sum_{i=1}^{M} r_{t,i}^2 + 2\sum_{h=1}^{q}(1 - \frac{h}{q+1})\sum_{i=1}^{M-h} r_{t,i}r_{t,i+h},$$

where q and h are natural numbers.

Corsi (2008) introduces a Heterogeneous Autoregressive model (HAR) for realized volatility and aims to capture some volatility features, such as the long-memory characteristic. The essence of this model lies in its logarithmic form, which is expressed as

$$v_t = b_0 + b_1 v_{t-1} + b_2 v_{t-5,t-1} + b_3 v_{t-22,t-1} + \varepsilon_t, \tag{5}$$

where $\varepsilon_t \sim D(0, \sigma^2)$ and $v_t = \log(RV_t)$. The variables $v_{t-5,t-1}$ and $v_{t-22,t-1}$ represent the weekly and monthly moving averages of log-realized volatilities, respectively, given by

$$v_{t-5,t-1} = \frac{\log(RV_{t-1}) + \log(RV_{t-2}) + \cdots + \log(RV_{t-5})}{5},$$

$$v_{t-22,t-1} = \frac{\log(RV_{t-1}) + \log(RV_{t-2}) + \cdots + \log(RV_{t-22})}{22}.$$

To incorporate the asymmetry and jumps into the volatility model, a more comprehensive version of the HAR model can be formulated as (Liu and Maheu, 2008):

$$\begin{aligned} v_t =& b_0 + b_1 v_{t-1} + b_2 v_{t-5,t-1} + b_3 v_{t-22,t-1} + b_J J_{t-1} + a_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \\ &+ a_2 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \mathbb{1}_{r_{t-1}<0} + \epsilon_t, \end{aligned} \tag{6}$$

where J_{t-1} represents a jump component given by

$$J_t = \begin{cases} \log(RV_t - RBP_t + 1), & \text{if } RV_t - RBP_t > 0, \\ 0, & \text{if } RV_t - RBP_t \leq 0, \end{cases}$$

and $RBP_t = \frac{\pi}{2} \sum_{i=1}^{M-1} |r_{t,i}||r_{t,i+1}|$ is termed as the Realized Bi-power Variation.

Later, we will compare both original HAR model and asymmetry HAR in this paper.

3. Cluster partition model

Volatility does not usually change over time in a uniform manner. Instead, it tends to cluster, resulting in periods where high volatility is followed by high volatility. Similarly, low volatility tends to be followed by low ones. This phenomenon is known as volatility clustering. An imperative aspect of this behavior is what's so-called a "volatility structural change". A discernible shift occurs, transiting the market from a period of high volatility to one of low volatility or vice-versa. Detecting such structural changes, especially the most recent ones, is of paramount importance for forecasting future volatility trends. This is because leveraging knowledge about the latest structural volatility can enhance the accuracy of such predictions. Thus, the primary challenge in this context becomes the identification of structural change points. The ultimate goal is to pinpoint the most recent change, serving as an indicator for impending volatility patterns.

Given a series of volatilities, denoted as $V = (\sigma_1, \ldots, \sigma_T)$, we wish to identify a partition point I that effectively divides V into distinct clusters. One method for detecting such partitioning is expressed by the following inequality[1]

$$\sum_{i=1}^{T} (\sigma_i - \bar{\sigma}_i)^2 > \sum_{i_1=1}^{I-1} (\sigma_{i_1} - \bar{\sigma}_{i_1})^2 + \sum_{i_2=I}^{T} (\sigma_{i_2} - \bar{\sigma}_{i_2})^2, \tag{7}$$

where the terms $\bar{\sigma}_i$, $\bar{\sigma}_{i_1}$ and $\bar{\sigma}_{i_2}$ represent the respective centroids of the full set and the two sub-sets divided by I. Often, the arithmetic or geometric mean values of the given subsets are used for these centroids.

[1]In fact, we can construct a statistic inference by Chi-square test for the series $\{v_i\}_{i=1}^{T}$ to search its partition point for two clusters. But here we adopt the simple form as the inequality (7) to be consistent with the latter Fisher's optimal dissection method.

The rationale is that if the difference between these two sides of inequality (7) is large, it indicates a significant difference between the two clusters. As the left-hand side is constant for a given dataset, our goal translates to minimizing the right-hand side. This leads to the cluster statistic $\rho_\sigma(I)$ for a single structural change:

$$\rho_\sigma(I) := \frac{1}{T}\sum_{i_1=1}^{I-1}(\sigma_{i_1} - \bar{\sigma}_{i_1})^2 + \frac{1}{T}\sum_{i_2=I}^{T}(\sigma_{i_2} - \bar{\sigma}_{i_2})^2. \tag{8}$$

The optimal division, $\arg\min_{1\le i\le T} \rho_\sigma(i)$, then is the one that minimizes ρ_σ across all potential partition points, ensuring that volatilities within each cluster are similar to each other and distinct from other clusters. This approach shares similarities with the cumulative sum (CUSUM) test applied to squared series, as discussed in Andreou and Ghysels (2002) and Xu (2013). Additionally, recognizing the possibility of multiple breakpoints, the same principle of minimizing deviations can be extended for detecting multiple structural changes.

3.1. General cluster partitions of dynamic volatility

Consider a classifier, denoted by $\pi(V, N)$. That means when applied to a volatility series V, this classifier divides V into N distinct clusters, represented as $\mathcal{C} = (C_1, \ldots, C_N)$. The divisions are determined by N split points $(I_1, I_2, \ldots, I_N)$, where $1 = I_1 < I_2 < \cdots < I_N < T$. For ease of notation, we introduce $I_{N+1} = T + 1$.

For each cluster C_k, the index I_k denotes its first starting point, with k ranging from 1 to N. It's crucial to note that out of the N split points in $\pi(V, N)$, only the initial point I_1 is predetermined. Thus, the classifier can be concisely represented as $\pi = (I_1, I_2, \ldots, I_N) \in \mathbb{N}_N^+$, which belongs to the positive natural numbers in ascending order.

Define a loss function $L(\pi)$ to evaluate the effectiveness of a given classifier, π. This function calculates the overall dispersion within the clusters formed by the classifier, given by

$$L(\pi) = \frac{1}{T}\sum_{k=1}^{N} D(V_{C_k}), \tag{9}$$

where $V_{C_k} = \{\sigma_i : i \in C_k\}$ is the collection of volatility within the k-th cluster, $D(V_{C_k}) = \sum_{i\in C_k} d(\sigma_i, \bar{\sigma}_{C_k})$ sums up the spread within the k-th cluster, comparing each volatility value to the cluster's center, $\bar{\sigma}_{C_k}$. The distance function $d(\cdot)$ measures

the gap between individual volatility value and the cluster's center. In this study, we're using the Euclidean distance as our measure. Specifically, it's calculated as $d(\sigma_i, \bar{\sigma}_{C_k}) = \|\sigma_i - \bar{\sigma}_{C_k}\|_2$.

Our goal is to minimize the value of this loss function. In other words, we're searching for a classifier that creates clusters with the least amount of internal dispersion. The best classifier is formally identified as

$$\pi^*(V, N) = \arg \min_{\pi \in \mathbb{N}_{N-1}^+} L\left[\pi(V, N)\right]. \tag{10}$$

By minimizing the loss function, we ensure that the created clusters have similar volatilities within them, making them as cohesive as possible.

The search for an optimal classifier can be viewed as an expansion of the binary classification problem shown in (8). Our goal is to find the optimal partition points within the volatility series V, resulting in the optimal clusters $(C_1^*, \ldots, C_N^*)$. While there are several methods available to identify the minimum value in (10), such as the Newton iteration, quasi-Newton method, steepest descent method, and conjugate gradient algorithm, they have their own sets of disadvantages. These may include dependencies on initial values, localization of the optimal solution, or slow rates of convergence. To address these challenges and seek a more comprehensive solution, we turn to artificial intelligence techniques. Specifically, we employ the genetic algorithm, a powerful tool renowned for its ability to approach global optima without being overly influenced by the aforementioned limitations. In this methodology, the optimal split points are denoted as $(I_1^*, \ldots, I_N^*)$. Notably, the classifier π^* is influenced solely by the conditional volatility V and the specified cluster number N. The optimality of this classifier arises from its ability to maintain the homogeneity of conditional volatility within individual clusters, while ensuring discernible differences across clusters. This characteristic aptly captures the nature of volatility clustering.

To ensure homogeneity within each cluster k, we can use a constant, $\overline{\sigma}_k$, as a representative volatility. The partitioned volatility, denoted as $V^{CP} = (\sigma_1^{CP}, \sigma_2^{CP}, \ldots, \sigma_T^{CP})$, can be articulated as

$$\sigma_t^{CP} = \mathbb{1}_{\{t \in C_k^*\}} \overline{\sigma}_k. \tag{11}$$

Considering the homogeneity of conditional volatility $\{\sigma_t^{CP} : t \in C_k^*\}$ within each cluster,

the value of $\overline{\sigma}_k$ is determined as the time-invariant arithmetic average of the associated cluster $\overline{\sigma}_k = \frac{1}{|C_k|}\sum_{i\in C_k}\sigma_i$. While there exist other potential forms for the constant representative volatility, we opt for the most straightforward method mentioned above.

Given the inherent homogeneity, the partitioned cluster volatility for the final cluster can serve as a predictor for future volatility

$$\sigma_{T+1} = \overline{\sigma}_N.$$

In subsequent sections, we will explore relaxing the constant nature of this setting, allowing it to be time-dependent within each cluster. For stationary time series, this revised approach ensures that the representative volatility remains consistent.

3.2. Iterated cluster partition volatility

Within each optimal cluster C_k^*, representative volatility doesn't necessarily need to be constant; it can be dynamic and dependent on the underlying model. Some possible models that can be employed include GARCH models and HAR, among others.

Applying the chosen model to fit the volatility within each cluster and estimating the relevant parameters, we can represent the estimated volatility as $\hat{V}_{C_k^*} = \{\hat{\sigma}_t^* : t \in C_k^*\}$ with the parameter notation $\hat{\theta}_k$. Combining these estimations across all clusters, we derive

$$\hat{V} = (\hat{V}_{C_1^*}, \ldots, \hat{V}_{C_N^*})$$

with the parameter set $\hat{\theta}^\top = (\hat{\theta}_1^\top, \ldots, \hat{\theta}_N^\top)$.

Interestingly, the estimated volatility $\hat{V}$ might manifest an even more pronounced clustering effect compared with the original V. By applying our cluster partition method to $\hat{V}$, we can obtain another classifier. If this new classifier matches the original, it indicates that our estimated volatility is optimally clustered. If not, the process is repeated until an optimally clustered volatility is achieved. This results is named Iterated Cluster Partition volatility V^{ICP}.

Procedure for Iterated Clustering Partition is provided as follows:

Step 1. Fit the entire sample's volatility, V.

Step 2. Apply cluster partitioning to V yielding the optimal classifier π_0^* and its associated clusters $\mathcal{C}_0$.

Step 3. Fit the volatility within each cluster to obtain $\hat{V}$.

Step 4. Using cluster partitioning on $\hat{V}$ results in classifier π_1^* and clusters $\mathcal{C}_1$.

Step 5. Assess the similarity of π_0^* and π_1^*. If they match, then $V^{ICP} = \hat{V}$. If not, set $V = \hat{V}$ and return to Step 2.

This iterative approach can be employed with any volatility model to yield a unique clustered volatility representation. [2]

3.3. Determining optimal cluster number

The cluster number N significantly impacts the optimal classifier, thus making the optimal determination of cluster number is a vital aspect of our methodology.

Conventionally, a straightforward method is to test the null hypothesis N clusters against an alternative hypothesis $N + 1$ clusters. However, as outlined by Hamilton (2010), this approach encounters certain challenges. Specifically, under the null hypothesis, some model parameters may remain unidentified, leading to complications in the test. This issue arises because, under certain scenarios (such as when only a single cluster is valid for the entire sample), the maximum likelihood estimation does not converge to any definite population metric. This non-convergence implies that the likelihood ratio test might not follow the customary χ^2 distribution. To interpret a likelihood ratio statistic one instead needs to appeal to the methods of Hansen (1992) and Garcia (1998).

There exist alternative tests that do not rely on the likelihood ratio statistic, as discussed by Carrasco et al. (2014). In a different vein, Bayesian methods offer another way. For instance, one can determine N based on the value that maximizes the marginal likelihood (Liu and Maheu, 2008). Other Bayesian techniques involve leveraging Bayesian factors (Koop and Potter, 2009) or evaluating models based on their forecasting power (Hamilton and Susmel , 1994).

Nevertheless, these strategies often have limitations. For instance, they typically detect no more than 10 regimes. This constraint can be problematic, especially when

[2]Due to our objective function (10) being a piecewise quadratic function, this function has either global convexity or piecewise local convexity. Consequently, we can ensure the convergence of the algorithm by utilizing classical iterative methods, such as Newton method, steepest descent method, conjugate gradient method, etc. In this article, we employ genetic algorithm to find the global optimal solution. This algorithm can also ensure its convergence in convex functions.

dealing with real-world financial data where volatility can exhibit dramatic variations, especially during economic crises. To overcome these challenges and the need for flexibility, our approach gravitates towards a data-driven methodology to identify the most appropriate cluster number.

In the realm of model selection, the information criterion aims to balance model fit and model complexity. To this end, we introduce a novel information-based statistic that combines both these dimensions. Specifically, this criterion takes into account the optimized loss function $L\left[\pi\left(V, N\right)\right]$ and introduces a penalty for model complexity, given by

$$\psi(N) = \log(L\left[\pi\left(V, N\right)\right]) + \frac{N\log(T)}{T}, \tag{12}$$

where the term $\frac{N\log(T)}{T}$ is the penalty for complexity. The logic behind this penalty is rooted in the principle that as the number of clusters increases, the risk of over-fitting also rises. Hence, the penalty discourages excessive clustering by penalizing the addition of each cluster, especially as the sample size T grows.

The objective now is to find the cluster number N that minimizes the information-based statistic. Formally, the optimal cluster number is given by

$$N^* = \arg\min_N \psi(N).$$

It is worth noting that the optimal cluster number is contingent upon the volatility series V and needs to be recalculated before each forecast. While this approach might increase computational effort, it promises improved forecasting accuracy in comparison to a static, fixed cluster number. By adapting the cluster number based on the evolving data characteristics, the methodology remains agile and responsive to the underlying data dynamics, thus enhancing its predictive power.

4. Comparison of models

Before presenting our empirical results, we summarize the key components of our theoretical model, cluster partition (CP) model, and how it sets the stage for the analysis that follows. This model segment volatility into distinct, non-overlapping clusters, enabling swiftly identifying of structural changes and leveraging the recent cluster for future volatility forecasting. The optimal cluster number is determined using an informa-

tion ratio. Through an iterative process of partition refinement and dynamic adjusting, we propose the iterated cluster partition (ICP) model.

The proposed CP and ICP models significantly advance beyond traditional GARCH or HAR models by capturing the structural clustering of volatility. Unlike these classic models, relying on a fixed form and predetermined lags, CP and ICP adapt dynamically to fitting market conditions. This adaptability is pivotal, especially during periods of high market volatility, enhancing the model's responsiveness and accuracy in forecasting.

GARCH models, constrained by fixed lags, often falter under rapidly changing market dynamics. During turbulent times shorter lags would be more informative whereas in more stable periods, longer lags might provide deeper insights. Our CP and ICP models circumvent this rigidity by actively detecting structural changes with real-time market dynamics. Specifically, during volatile market phases, these models shorten the last cluster's length to reflect recent, and extend it during more stable periods to provide a more comprehensive view. Similarly, HAR models, utilizing fixed lags at different frequencies, suffer from a rigid window for estimating coefficients of HAR, potentially impairing forecast accuracy. In contrast, our models offer the flexibility to dynamically estimate and forecast volatility, enhancing predictive performance across various market conditions.

To bring out the distinctiveness and potential advantages of the volatility cluster partition model, it's crucial to situate it in the context of existing volatility models. Before leveraging our cluster partition methodology on the volatility series V, we'll evaluate the performance of cluster partition and iterated cluster partition models. This comparative exercise will help elucidate the nuances and effectiveness of our model comparing with traditional ones. The array of models we consider for this comparative results are outlined in Table 1. Each model captures the volatility dynamics differently, and by juxtaposing them against the cluster partition model, we aim to highlight the uniqueness of our approach. The real test of a model's efficiency lies in its performance in practical scenarios. Hence, to showcase the potential superior performance of our model, we undertake a two-pronged comparative analysis: volatility forecasting and VaR estimation. For each model, we forecast volatility for a set horizon using a rolling-window approach. The forecast accuracy is evaluated using various metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Using the forecasted volatility, we

estimate the VaR for different confidence intervals, 90%, 95% and 99%. This allows us to assess how well the model captures extreme tail risks, which is essential for risk management and regulatory purposes.

Table 1: Model summary

Basic model	**Cluster method**	**Short**
GARCH	no clustering	GARCH
	cluster partition	CPGARCH
	iterative cluster partition	ICPGARCH
GJR-GARCH	no clustering	GJR
	cluster partition	CPGJR
	iterative cluster partition	ICPGJR
HAR	no clustering	HAR
	cluster partition	CPHAR
	iterative cluster partition	ICPHAR
HAR-a	no clustering	HAR-a
	cluster partition	CPHAR-a
	iterative cluster partition	ICPHAR-a

Note: This table presents the summary of used models and the corresponding short letters.

In our comparative analysis, we adopt a moving-window approach to evaluate the forecasting performance of volatility models. We opt for an estimation window spanning $T = 750$ days, which is roughly equivalent to the average number of trading days over three years. We evaluate models across three distinct adjustment frequencies: daily, weekly and monthly. For the sake of clarity and brevity, we primarily focus on the results derived for a 1-period ahead forecast. Other frequencies give similar results and could be provided by authors on request.

For every forecast period t, utilizing data from the preceding days, the parameters necessary for a specific model are estimated. These models are subsequently employed to project the volatility σ_t. Leveraging this forecasting volatility, we then compute VaR_t. This procedure is iteratively executed: the time series is expanded by appending one daily returns for the succeeding period, whilst concurrently omitting the earliest returns. This progress continues until the end of data-set.

4.1. Volatility forecast

Volatility forecast evaluations play a vital role in financial studies, underscoring the importance of accurate forecasts in informed decision-making processes. Evaluating the efficiency of one model over another requires an assessment of its forecasting accuracy.

Our primary evaluation metrics for out-of-sample forecasting performance are two loss functions, MAE and RMSE:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\sigma_{a,t} - \sigma_{f,t}|,$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}|\sigma_{a,t} - \sigma_{f,t}|^2},$$

where n signifies the total number of forecasts, and $\sigma_{a,t}$ and $\sigma_{f,t}$ denote the true and forecasted volatilities, respectively.

It's essential to understand that volatility, being intrinsically a latent, remains unobservable. Yet, there exists several proxies deemed as unbiased observations for volatility, such as squared returns, range, and realised volatility. Among them, realised volatility stands out for its robustness (Patton, A.J., 2011). Therefore, we employ realised volatility as our chosen proxy for the true volatility. To supplement loss functions, we apply the DM test (Diebold and Marriano, 1995), to compare the forecast performance of models in pairs. The DM test's null hypothesis suggests that two competing forecasts yield identical errors. The amalgamation of the rolling forecasting methodology and the concluding DM test enables us to discern the model that exhibits the best performance.

4.2. VaR estimation

Value-at-Risk (VaR) is very important in financial risk management, because of its straightforward representation and its easily interpretable nature. Its wide-ranging applications span various domains, including investment portfolios (Chen et al., 2019) and commodity markets (Chkili, 2014). Defined for a specific probability level α, VaR represents the α-quantile of the conditional distribution of asset returns. Its prominence grew exponentially since 1994 following J.P. Morgan's introduction of the Risk-Metrics system. Subsequently, in 1996, the Basel Committee on Banking Supervision advocated that banks should use internal VaR models to determine the requisite capital backing

their trading operations.

Consider $P(r_t \leq x)$ as the probability that the asset return does not exceed the value x. We can denote the cumulative distribution function (CDF) of r_t as $F(x) = P(r_t \leq x)$. For a predetermined significance level α $(0 < \alpha < 1)$, the VaR can be mathematically defined as:

$$\mathrm{VaR}_t(\alpha) = -\sup\left\{x : F(x) \leq \alpha\right\}. \tag{13}$$

To estimate VaR, we predominantly utilize two methodologies: the historical simulation approach and the model-based mean variance method, as detailed by Chen et al. (2019).

The historical simulation approach operates on the premise that historical returns are indicative of future performance. The empirical CDF of asset returns is defined as

$$\widehat{F}^h(x) = \frac{1}{n}\sum_{\tau=t-n}^{t-1} \mathbb{1}\left(r_\tau < x\right), \tag{14}$$

where n represents the window size, typically selected as 125, 250 or 250. These window sizes correspond respectively to half a year, a full year and two years of daily observations. Consequently, the VaR forecast is

$$\widehat{\mathrm{VaR}}_t(\alpha) = -\sup\left\{x : \widehat{F}^h(x) \leq \alpha\right\}. \tag{15}$$

The model-based mean variance approach employs the ARMA-GARCH dynamics for computing the conditional mean and variance, as depicted in equation (1). The VaR, in this context, is calculated using the CDF F_ε as

$$\widehat{\mathrm{VaR}}_t(\alpha) = \mu_t + \sigma_t F_\varepsilon^{-1}(\alpha). \tag{16}$$

Indetermining the distribution of ε_t, both normal and student t distributions are frequently employed.

In our analysis, we calculate VaRs across a spectrum of pre-defined significance levels, ranging from 1% to 10%. These computations are subsequently assessed using two established backtesting methodologies: the unconditional coverage backtest (Kupiec, 1995), and the conditional coverage backtest (Engle and Manganelli, 2004).

To exam the failure rate, we employ the Kupiec Likelihood Ratio (LR) test. The statistic for the Kupiec LR test is

$$LR = -2\ln\left[(1-\alpha)^{T-N}\alpha^{N}\right] + 2\ln\left[(1-f)^{T-N}f^{N}\right] \tag{17}$$

where N denotes the count of return observations surpassing the estimated VaR, and T represents the sample size. Under the null hypothesis, which posits that the realized failure rate matches the stipulated confidence level α, the Kupiec LR statistic adheres to an asymptotic chi-squared distribution with one degree of freedom.

The conditional coverage of our backtest is conducted via the Dynamic Quantile (DQ) backtest. The model underpinning this approach is formulated as

$$H_{t+1} - \alpha = \gamma_0 + \gamma_1 H_t(\alpha) + \gamma_2 \mathrm{VaR}_{t+1}(\alpha) + \epsilon_{t+1}$$

where $H_t = \mathbb{1}\left(r_\tau < \mathrm{VaR}_t\right) - \alpha$.

The choice of regressors in our model is inspired by Berkowitz et al. (2011). To validate the efficacy of our conditional coverage backtest, we implement the Wald test, examining the null hypothesis that γ_0, γ_1 and γ_2 are jointly equal to zero. The test statistic for this hypothesis adheres to an asymptotic χ^2_3 distribution.

5. Empirical results

We collect daily close prices of S&P 500 index, DAX 30 of German stocks, and FTSE 100 index of UK stocks, with the dates spanning from May 18, 2012 to May 19, 2022. The data is obtained from the Bloomberg database.

Let P_t denote the index price on day t. Returns are computed using log returns, $r_t = \ln(P_t/P_{t-1})$. Table 2 offers a comprehensive overview of the summary statistics derived from the full-sample return series. The range of average annualized returns is between 2.244% for FTSE and 6.632% for the S&P 500. Correspondingly, annualized standard deviations vary between 16.476% and 20.805%. Every return series showcases a negative skewness (approximately -0.6) and high kurtosis (approximately 13.6). While the auto-correlations of the return series hover around 0, the auto-correlations of squared returns are all positive. The positive squared correlation means that there exists volatility clustering, which is consistent with literatures.

Table 2: Summary statistics of S&P 500, DAX, and FTSE 100 index

	Mean (%)	**Std (%)**	**Skewness**	**Kurtosis**	**Corr**(r)	**Corr**(r^2)
S&P 500	6.632	20.805	-0.555	15.701	-0.140	0.343
DAX	5.736	20.748	-0.500	10.981	0.014	0.176
FTSE 100	2.244	16.476	-0.744	13.587	-0.006	0.268

Note: This table presents summary statistics on the daily equity return series, over the full sample period from 18 May 2012 to 18 May 2022. The first two rows report the annualized mean and standard deviation of these returns in percent. Corr is the auto-correlation corresponding to returns and squared returns. All these three indices reflect strong ARCH effect.

Following the approach of Andersen et al. (2006), who posited that realized volatility serves as an effective proxy for actual volatility, we also sourced 5-minute intra-day data for the S&P 500, DAX 30, and FTSE 100 from Bloomberg.

5.1. In-sample estimations of volatility models

In Table 3, we show the results of estimated parameters for the standard GARCH and GJR-GARCH model over the in-sample period spanning from May 18, 2012 to May 18, 2016. The upper panel shows the parameters of the GARCH(1,1) model, while the lower panel presents those of GJR-GARCH(1,1) model. The standard deviations are reported in the parentheses. For the conditional variance estimation, we observe that the stationary condition is satisfied for all of three indices since the sum of the ARCH and GARCH coefficients is less than one. Almost all coefficients are significant except for ARCH terms in GJR-GARCH. However, the additional leverage term is significantly positive. In particular, the GARCH coefficients for both GARCH and GJR-GARCH model are generally higher than 0.7, which indicates high volatility persistence in all cases.

Table 3: Parameters estimation of GARCH and GJR-GARCH model

	SP500	**DAX**	**FTSE**
w	0.00***	0.00*	0.00***
	(0.00)	(0.00)	(0.00)
α	0.17***	0.09***	0.12***
	(0.04)	(0.02)	(0.03)
β	0.71***	0.87***	0.80***
	(0.06)	(0.04)	(0.03)
w	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)
α	0.00	0.00	0.00
	(0.03)	(0.02)	(0.02)
β	0.76***	0.85***	0.80***
	(0.04)	(0.02)	(0.03)
γ	0.33***	0.18***	0.26***
	(0.06)	(0.04)	(0.05)

Note: $*$, $**$ and $***$ indicate statistical significance at the 10%, 5%, and 1% levels, respectively. In the upper and lower panels, the parameters are estimated by GARCH and GJR-GARCH model, respectively.

Similarly, in Table 4, the upper panel displays the estimated parameters of the HAR model, and the lower panel lists out the detailed ones of the HAR-a model. These parameters indicate RV is daily and weekly autocorrelated for all indices, but not significantly monthly autocorrelated. Specifically, leverage effect is significant in HAR-a model.

Table 4: Parameter estimations by HAR and HAR-a model

	SP500	**DAX**	**FTSE**
b_0	-2.78***	-1.72***	-1.66***
	(0.58)	(0.45)	(0.42)
b_1	0.43***	0.24***	0.28***
	(0.04)	(0.04)	(0.04)
b_2	0.29***	0.52***	0.48***
	(0.06)	(0.07)	(0.07)
b_3	0.02	0.06	0.08
	(0.07)	(0.07)	(0.06)
b_0	-2.82***	-1.68***	-1.63***
	(0.56)	(0.44)	(0.42)
b_1	0.33***	0.13***	0.19***
	(0.05)	(0.05)	(0.05)
b_2	0.36***	0.62***	0.55***
	(0.07)	(0.58)	(0.07)
b_3	0.05	0.08	0.10*
	(0.07)	(0.07)	(0.06)
b_J	77.19	723.95	-1841.50
	(71.70)	(672.53)	(1728.80)
a_1	0.00	-0.06**	-0.02
	(0.02)	(0.03)	(0.02)
a_2	0.22***	0.24***	0.15***
	(0.04)	(0.04)	(0.03)

Note: In the upper and lower panels, the parameters are estimated by HAR and HAR-a model, respectively. In this table, $*$, $**$ and $***$ indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

To elucidate our approach to the cluster partition method, we use the S&P 500 index as a representative example for our analysis. Panels 1-4 of Figure 1 display different volatility series for the S&P 500 index by GARCH, GJR-GARCH, HAR, and HAR-a models. Analogous results for other indices can be seen in Figures B.10 and B.11, respectively.

For each volatility series generated, the cluster partition method is employed to discern distinct clusters. To ensure the robustness of our findings, we randomly choose one of the optimal cluster numbers. For instance, from the HAR model the selected

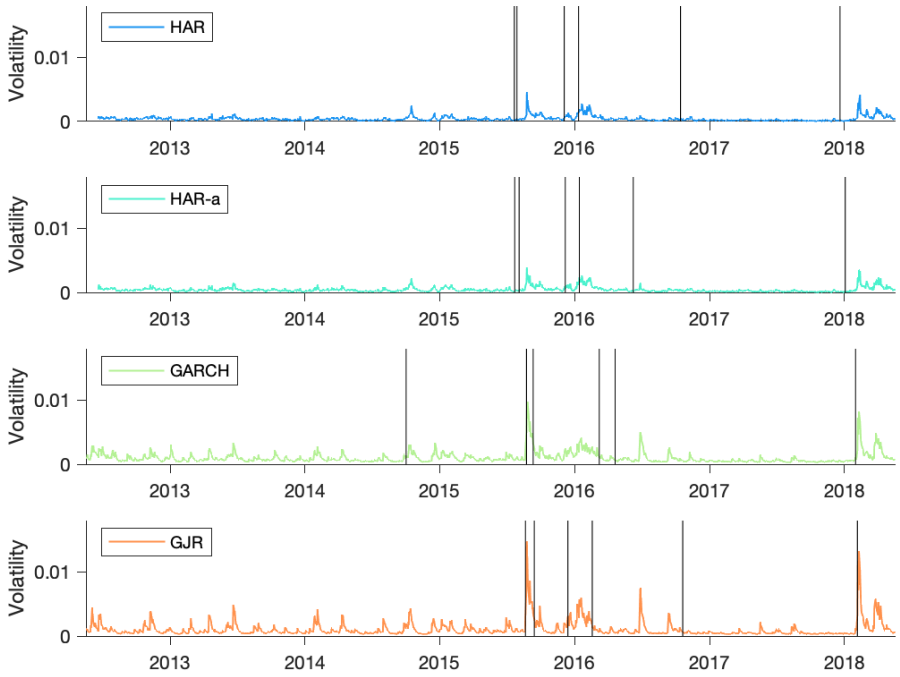

Figure 1: Cluster Partition Volatility of S&P 500 index. In Panel 1-4, we show conditional volatility estimated by GARCH, GJR-GARCH, HAR and HAR-a. For each volatility series, we use cluster partition method to get different clusters.

optimal cluster numbers are used to other models. As illustrated in Figure 1, each volatility series comprises seven clusters. The specific locations of split points diverge across each volatility sequence, yet we discern similar overarching patterns. A majority of the cluster split points are located at post-2015, signaling that volatility prior to this year was comparably stable. For every volatility series, the final split point emerges around 2018, suggesting a shift into a different volatility cluster after this year.

5.2. Volatility forecast

We now shift focus to the out-of-sample volatility forecasting performance. To compare the effectiveness of these models, we conduct out-of-sample volatility forecasts employing a rolling window methodology. The out-of-sample period spans from May 18, 2018 to May 18, 2022, encompassing 1008 time points. For each specific time point t within the forecast set, models are estimated using a fixed window size of 750 observations. To check the robustness of our findings, we also exercise with different window sizes as 500 and 1000; the results, which align closely with the primary analysis, are listed out in detail in Appendix B. Leveraging these estimated outcomes, we then predict the next day's volatility using various models, including the original models, the Cluster Partition (CP) model and the Iterated Cluster partition (ICP) model.

The outcomes of volatility forecasting across various methodologies are presented in detail in Table 5. In the upper panel, we observe that CP and ICP methods, when applied to S&P 500 index, do not consistently enhance forecasting accuracy over GARCH and GJR models in terms of MAE and RMSE for the entire out-of-sample period. Notably, CPGARCH, ICPGARCH, CPGJR and ICPGJR models perform poorly in 2018-19 and 2020-21, exhibiting higher MAE and RMSE. However, during 2019-20 and 2021-22, these models show improvements over the original counterparts.

The same pattern can be also found in HAR and HAR-a and their extensions in the lower panel of Table 5. Given that HAR models are based on realized volatility, a direct comparison with models relying solely on the daily returns (upper panel) is not straightforward. Nevertheless, during the periods 2019-20 and 2021-22, both CPHAR and ICPHAR models considerably outperform the basic HAR model. Meanwhile, CPHAR-a and ICPHAR-a also considerably surpass HAR-a.

The CP and ICP models particularly excel during the highly volatile periods of 2019-20 and 2021-22, which include the Covid-19 market disruptions in March 2020. This

suggests that CP and ICP models are more adept at handling periods of significant market fluctuations.

Table 5: Comparison of S&P 500 index volatility forecast performance [%] across competing models with estimation window 750 days

	2018-19		**2019-20**		**2020-21**		**2021-22**	
	MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
GARCH	4.17	5.28	9.01	14.33	**6.21**	**7.72**	5.76	6.85
CP-GARCH	5.43	6.83	**8.56**	**14.05**	7.47	8.94	**5.38**	**6.33**
ICP-GARCH	5.48	6.90	**8.56**	14.16	7.47	8.94	**5.38**	**6.33**
GJR	**4.10**	**5.25**	8.70	15.13	6.29	8.33	5.99	7.17
CP-GJR	6.10	7.48	8.70	16.86	7.36	9.82	5.45	6.40
ICP-GJR	5.98	7.34	8.75	17.15	7.38	9.83	5.45	6.40
HAR	2.83	4.26	4.38	8.27	4.03	5.51	3.26	**4.55**
CP-HAR	2.96	4.34	4.36	8.20	4.03	5.48	**3.25**	4.56
ICP-HAR	2.97	4.36	4.37	8.20	4.03	5.48	**3.25**	4.58
HAR-a	**2.78**	**4.18**	4.38	8.24	**3.96**	5.40	3.27	**4.55**
CP-HAR-a	2.91	4.26	**4.35**	**8.18**	3.97	**5.35**	3.33	4.64
ICP-HAR-a	2.90	4.26	**4.35**	**8.18**	3.97	**5.35**	3.31	4.64

Notes: This table reports the mean losses of the different volatility models over the out-of-sample period with respect to two MAE and RMSE. The values in bold face indicate best-performing models (i.e. models with the lowest mean losses).

Table 6 presents Diebold-Mariano t-statistics for loss differences concerning the S&P 500 index. The test, conducted as the "row model minus column model", indicates superior performance for a column model over a row model, with positive value. Statistics significant above 95% confidence level are highlighted in red. Panel A shows the test results of 2018-2022. Notably, HAR extensions consistently outperform GARCH family models, demonstrating significantly positive values for GARCH family rows. Specifically, the entirety of entries in the ICPHAR-a column are positive, unequivocally confirming its superior forecasting accuracy. In addition, CPHAR and ICPHAR present positive values against standard HAR, and CPHAR-a exhibits positive values compared to HAR-a. While these improvements are not statistically significant, they suggest that CP and ICP methods could somehow enhance the accuracy of the established volatility forecast

models. However, these methods do not show improvements for GARCH and GJR over the entire period.

Panel B shows the specific test results for period 2019-2020, as we already discussed, this period contains high volatility fluctuation, which reflects the challenges of volatility forecasting. ICPHAR-a still remains as the top-performing model. Moreover, CP and ICP methods enhance the forecasts of HAR and HAR-a models. Interestingly, we find that CPGARCH and ICPGARCH show positive results compared with standard GARCH, and CPGJR and ICPGJR are positive relative to GJR. This improvement during a highly volatile period aligns with findings from Table 5, and confirms that CP and ICP models are particularly effective in turbulent market conditions. Similar patterns are observed for other indices as details in Appendix B.

Figure 2 presents the out-of-sample volatility forecasts of GARCH, GJR, HAR, HAR-a and their extensions. Realised volatility, indicated by the red lines, serve as a proxy for true volatility. GARCH and GJR models under-perform, as they deviate from realised volatility. This discrepancy largely stems from their reliance solely on daily price, whereas both HAR and HAR-a models integrate intra-day prices, providing a more comprehensive model of volatility. Notably, extensions of GARCH and GJR particularly struggle during 2019-2020 period.

To ensure the robustness of our findings, we also modify the estimation window size to either 500 or 1000, as detailed in Tables B.9 and B.10 in Appendix B.

5.3. VaR forecast

For a comparative analysis of VaR results derived from various methods, we primarily concentrate on a single index. The out-of-sample time period is the same as that in volatility forecasting. We have also explored other indices, and obtained the similar results, which are listed out in detail in Appendix B. Notably, the historical simulation serves as our benchmark method.

Figure 3-7 display the daily returns of the index along with the one-day-ahead forecast confidence intervals produced by various models at distinct significance levels. The derivation of these confidence intervals is based on the assumption that the errors either adhere to a normal distribution or a skewed-t distribution. Subsequent to this assumption, we estimate the relevant distribution parameters using the corresponding index.

Table 6: DM t-statistics on average out-of-sample loss differences, S&P 500 volatility

Panel A: 2018-2022

	GARCH	**CPGARCH**	**ICPGARCH**	**GJR**	**CPGJR**	**ICPGJR**	**HAR**	**CPHAR**	**ICPHAR**	**HAR-a**	**CPHAR-a**	**ICPHAR-a**
GARCH		-1.06	-1.19	-1.38	-1.83	-1.83	4.98	5.17	5.16	5.05	5.17	5.17
CPGARCH	1.06		-0.98	-0.46	-1.93	-1.90	4.49	4.50	4.49	4.55	4.50	4.5
ICPGARCH	1.19	0.98		-0.32	-1.90	-1.89	4.57	4.55	4.55	4.61	4.55	4.55
GJR	1.38	0.46	0.32		-1.79	-1.80	4.26	4.34	4.33	4.31	4.36	4.36
CPGJR	1.83	1.93	1.90	1.79		-0.64	3.45	3.43	3.43	3.48	3.44	3.44
ICPGJR	1.83	1.90	1.89	1.80	0.64		3.43	3.41	3.40	3.46	3.41	3.41
HAR	-4.98	-4.49	-4.57	-4.26	-3.45	-3.43		0.11	0.06	0.85	0.28	0.29
CPHAR	-5.17	-4.50	-4.55	-4.34	-3.43	-3.41	-0.11		-1.39	0.23	0.43	0.44
ICPHAR	-5.16	-4.49	-4.55	-4.33	-3.43	-3.40	-0.06	1.39		0.28	0.51	0.53
HAR-a	-5.05	-4.55	-4.61	-4.31	-3.48	-3.46	-0.85	-0.23	-0.28		0.01	0.02
CPHAR-a	-5.17	-4.50	-4.55	-4.36	-3.44	-3.41	-0.28	-0.43	-0.51	-0.01		0.15
ICPHAR-a	-5.17	-4.50	-4.55	-4.36	-3.44	-3.41	-0.29	-0.44	-0.53	-0.02	-0.15	

Panel B: 2019-2020

	GARCH	**CPGARCH**	**ICPGARCH**	**GJR**	**CPGJR**	**ICPGJR**	**HAR**	**CPHAR**	**ICPHAR**	**HAR-a**	**CPHAR-a**	**ICPHAR-a**
GARCH		0.35	0.21	-0.88	-1.08	-1.14	3.48	3.64	3.64	3.50	3.61	3.61
CPGARCH	-0.35		-0.76	-1.17	-1.53	-1.56	2.61	2.63	2.63	2.63	2.61	2.62
ICPGARCH	-0.21	0.76		-1.07	-1.51	-1.56	2.67	2.68	2.68	2.68	2.66	2.66
GJR	0.88	1.17	1.07		-1.02	-1.11	2.92	2.99	2.99	2.94	2.98	2.99
CPGJR	1.08	1.53	1.51	1.02		-0.79	2.24	2.23	2.23	2.25	2.23	2.23
ICPGJR	1.14	1.56	1.56	1.11	0.79		2.28	2.26	2.26	2.28	2.26	2.26
HAR	-3.48	-2.61	-2.67	-2.92	-2.24	-2.28		0.16	0.17	0.16	0.17	0.18
CPHAR	-3.64	-2.63	-2.68	-2.99	-2.23	-2.26	-0.16		1.06	-0.11	0.08	0.11
ICPHAR	-3.64	-2.63	-2.68	-2.99	-2.23	-2.26	-0.17	-1.06		-0.12	0.06	0.09
HAR-a	-3.50	-2.63	-2.68	-2.94	-2.25	-2.28	-0.16	0.11	0.12		0.14	0.16
CPHAR-a	-3.61	-2.61	-2.66	-2.98	-2.23	-2.26	-0.17	-0.08	-0.06	-0.14		0.79
ICPHAR-a	-3.61	-2.62	-2.66	-2.99	-2.23	-2.26	-0.18	-0.11	-0.09	-0.16	-0.79	

Notes: This table presents t-statistics from DM tests comparing the average loss, during the out-of-sample period from May 18, 2018 to May 18, 2022 across different forecast models. A positive value indicates that the column model outperform the row model in terms of average forecast accuracy. Statistics significant at the 95% confidence level are highlighted in red.

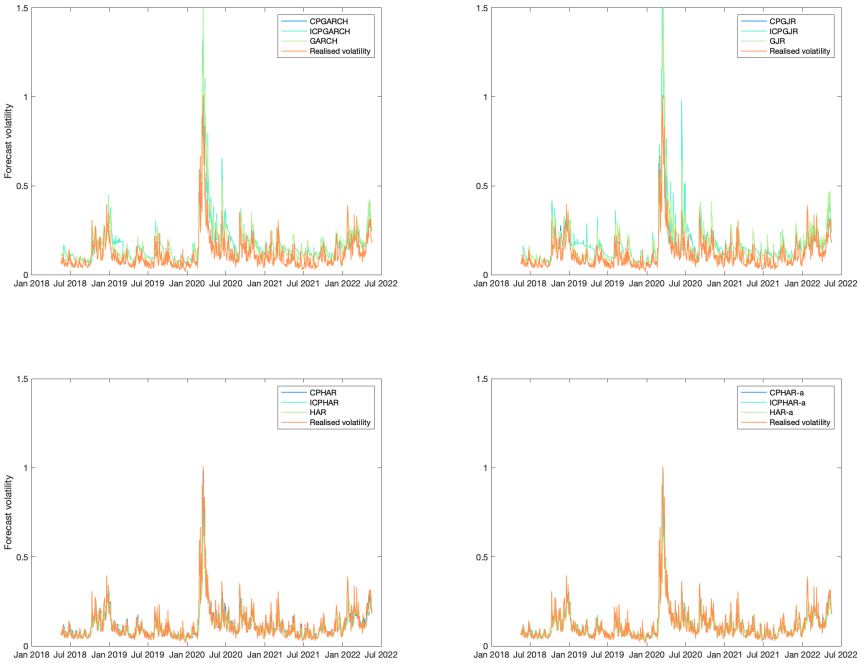

Figure 2: Volatility forecast by GARCH, GJR, HAR, HAR-a and their extensions.

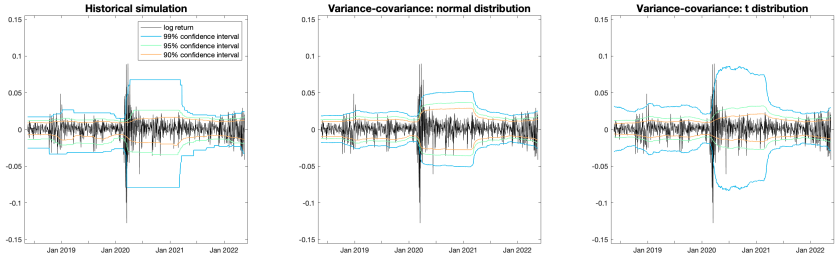

Figure 3: S&P 500 log returns confidence intervals of historical simulation, variance-covariance normal distribution and variance-covariance t distribution.

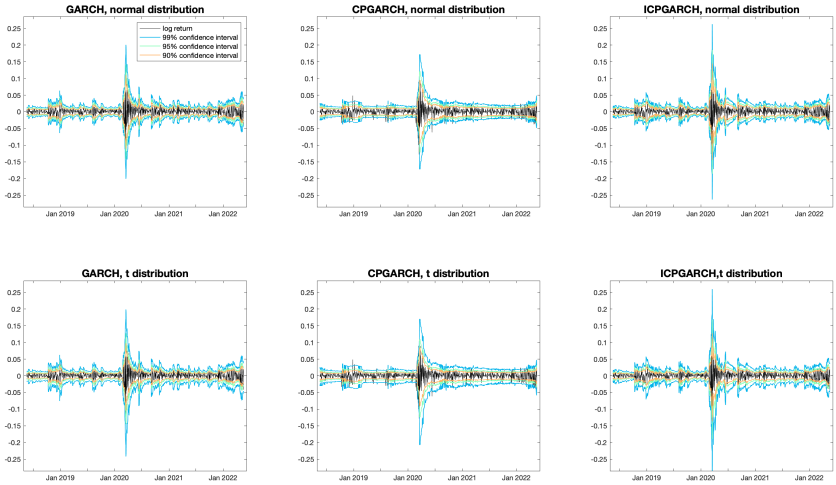

Figure 4: S&P 500 log returns confidence intervals of GARCH, CPGARCH and ICPGARCH. Innovations are either normal or skewed-t distribution.

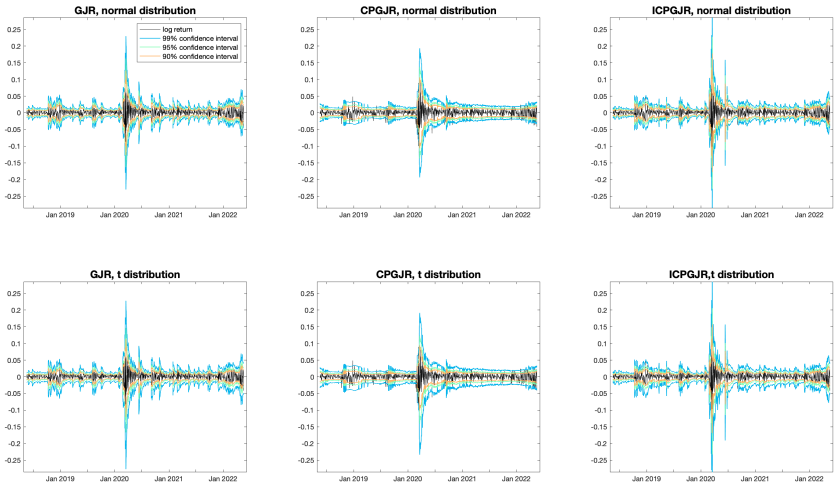

Figure 5: S&P 500 log returns confidence intervals of GJR, CPGJR and ICPGJR. Innovations are either normal or skewed-t distribution.

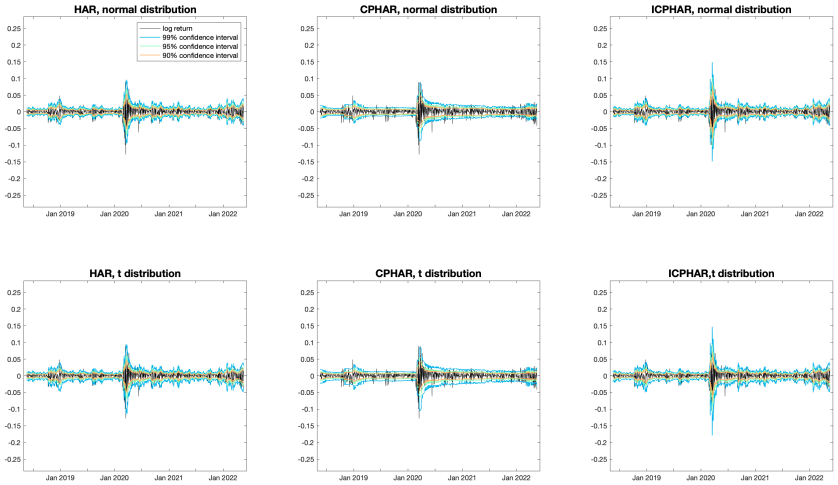

Figure 6: S&P 500 log returns confidence intervals of HAR, CPHAR and ICPHAR. Innovations are either normal or skewed-t distribution.

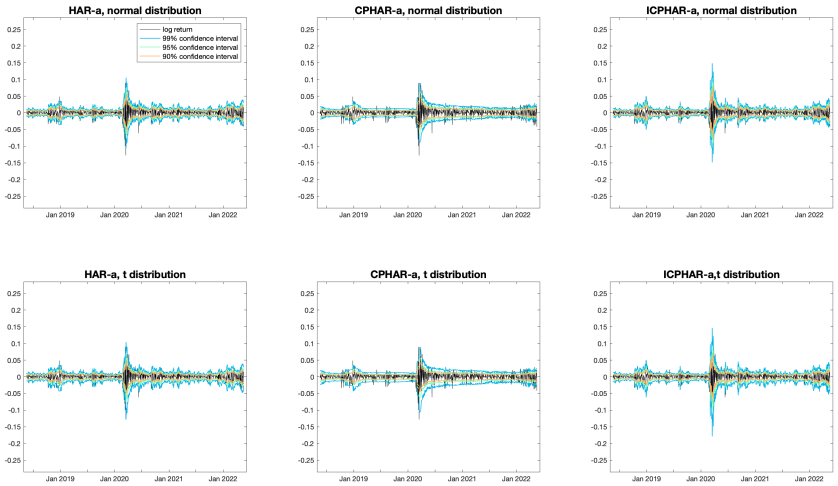

Figure 7: S&P 500 log returns confidence intervals of HAR-a, CPHAR-a and ICPHAR-a. Innovations are either normal or skewed-t distribution.

Table 7 presents the Kupiec unconditional tests for the returns of S&P 500 index at a confidence level of 95%. The first column shows the p-values of the entire 4-year out-of-sample from 2018 to 2022. Models utilizing a heavy-tailed skewed-t distribution consistently outperform those with a normal distribution. Specifically, GARCH models surpass HAR models in terms of VaR, largely because HAR models tend to forecast relatively lower volatility and overestimate risk. This lower volatility results in a higher incidence of VaR violation.

Furthermore, CP and ICP models enhance the VaR estimation capabilities of both GARCH and GJR models, irrespective with of innovation distribution used. This improvement is especially notable during the high volatility period from 2019 to 2020, underscoring the ability of CP and ICP methods to integrate recent data effectively and estimate more accurate VaR in extreme conditions. However, it is important to acknowledge that CP and ICP models have limitations. During the period from 2021 to 2022, these models exhibit less effective VaR estimation, indicating that while CP and ICP models are robust under certain conditions, they may not consistently outperform across all market phases.

Table 7: Kupiec unconditional tests of S&P under 95% confidence level

	2018-2022	**2018-2019**	**2019-2020**	**2020-2021**	**2021-2022**
Historical Simulation	**0.1375**	**0.4906**	0.0260	0.0038	0.0283
GARCH-N	0.0430	**0.6799**	0.0479	**0.9199**	0.1499
GARCH-t	**0.1375**	**0.8969**	0.0843	**0.9199**	**0.3603**
CPGARCH-N	*0.0791*	**0.4906**	**0.1414**	**0.6267**	*0.0900*
CPGARCH-t	**0.2257**	**0.8969**	**0.2261**	**0.6267**	**0.1499**
ICPGARCH-N	*0.0791*	**0.4906**	**0.1414**	**0.6267**	*0.0900*
ICPGARCH-t	**0.2257**	**0.8969**	**0.2261**	**0.6267**	**0.1499**
GJR-N	*0.0588*	**0.6799**	*0.0843*	**0.7016**	**0.2381**
GJR-t	**0.1375**	**0.6799**	**0.1414**	**0.9199**	**0.3603**
CPGJR-N	**0.1774**	**0.6468**	**0.1414**	**0.8500**	*0.0900*
CPGJR-t	**0.2257**	**0.4446**	**0.1414**	**0.8500**	*0.0900*
ICPGJR-N	**0.1774**	**0.6468**	**0.1414**	**0.8500**	*0.0900*
ICPGJR-t	**0.2257**	**0.4446**	**0.1414**	**0.8500**	*0.0900*
HAR-N	0.0000	0.0003	0.0000	0.0142	0.0000
HAR-t	0.0000	0.0003	0.0000	*0.0871*	0.0000
CPHAR-N	0.0000	0.0014	0.0000	**0.1456**	0.0000
CPHAR-t	0.0000	0.0030	0.0000	**0.2321**	0.0000
ICPHAR-N	0.0000	0.0014	0.0000	**0.1456**	0.0000
ICPHAR-t	0.0000	0.0030	0.0000	**0.2321**	0.0000
HAR-a-N	0.0000	0.0006	0.0000	0.0271	0.0000
HAR-a-t	0.0000	0.0006	0.0000	*0.0871*	0.0000
CPHAR-a-N	0.0000	0.0003	0.0000	**0.3524**	0.0000
CPHAR-a-t	0.0000	0.0014	0.0000	**0.7016**	0.0000
ICPHAR-a-N	0.0000	0.0003	0.0000	**0.3524**	0.0000
ICPHAR-a-t	0.0000	0.0014	0.0000	**0.5096**	0.0000

Notes: This table presents p-values from the Kupiec unconditional tests of VaR. Values that are greater than 0.1 (indicating no evidence against the corresponding level) are in bold, and values between 0.05 and 0.1 are in italics.

In addition, we present the results of conditional DQ tests for S&P 500 index returns at a confidence level of 95% in Table 8. The DQ test results corroborate that the CP and ICP methods enhance the VaR estimation performance of GARCH model across different innovation distributions. Although the GJR model with skewed-t distribution outperform the corresponding CPGJR and ICPGJR models, when using a normal distribution, both CPGJR and ICPGJR demonstrate superior performance over the standard

GJR model.

Specifically, focusing on the highly volatile period from 2019 to 2020, the CP and ICP models significantly improve the VaR estimation capabilities of the original GARCH and GJR models. This finding aligns with the results of the unconditional tests, reinforcing the robustness of the CP and ICP methods under extreme market conditions.

Table 8: DQ tests for S&P 500 under 95% condidence level

	2018-2022	**2018-2019**	**2019-2020**	**2020-2021**	**2021-2022**
Historical Simulation	0.0169	**0.5814**	0.1113	0.0000	**0.6141**
GARCH-N	**0.3626**	**0.9769**	**0.1965**	0.0000	0.0000
GARCH-t	**0.5911**	0.0000	**0.2140**	0.0000	0.0000
CPGARCH-N	**0.4084**	**0.4286**	**0.2685**	0.0000	0.0000
CPGARCH-t	**0.6668**	**0.4999**	**0.2568**	0.0000	0.0000
ICPGARCH-N	**0.4059**	**0.3680**	**0.2682**	0.0000	0.0000
ICPGARCH-t	**0.6634**	**0.4302**	**0.2558**	0.0000	0.0000
GJR-N	**0.3352**	0.0000	**0.1770**	0.0000	0.0000
GJR-t	**0.5613**	0.0000	**0.1562**	0.0000	0.0000
CPGJR-N	**0.3656**	0.0000	**0.2674**	0.0000	0.0000
CPGJR-t	**0.4177**	0.0000	**0.2674**	0.0000	0.0000
ICPGJR-N	**0.3642**	0.0000	**0.2653**	0.0000	0.0000
ICPGJR-t	**0.4160**	0.0000	**0.2653**	0.0000	0.0000
HAR-N	0.0000	0.0000	0.0001	*0.0858*	0.0000
HAR-t	0.0000	0.0000	0.0001	0.0000	0.0000
CPHAR-N	0.0003	0.0045	0.0001	**0.2256**	0.0000
CPHAR-t	0.0003	0.0242	0.0001	0.0000	0.0001
ICPHAR-N	0.0000	0.0046	0.0001	**0.2256**	0.0000
ICPHAR-t	0.0000	0.0241	0.0001	0.0000	0.0001
HAR-a-N	0.0000	0.0008	0.0001	**0.1668**	0.0001
HAR-a-t	0.0000	0.0008	0.0001	**0.3170**	0.0001
CPHAR-a-N	0.0006	0.0040	0.0001	0.0000	0.0000
CPHAR-a-t	0.0000	0.0191	0.0001	0.0000	0.0000
ICPHAR-a-N	0.0000	0.0040	0.0001	0.0000	0.0000
ICPHAR-a-t	0.0000	0.0183	0.0001	0.0000	0.0000

Notes: This table reports the DQ test p-values of the different volatility forecasting models. Values greater than 0.1 (indicating no evidence against the corresponding level) are in bold, and values between 0.05 and 0.1 are in italics.

6. Conclusion

This paper considers the concept of volatility clustering and introduces a novel volatility cluster partition model to understand and predict stock market volatility through a unique lens. This methodology aims to segment volatility into distinct clusters based on Fisher's optimal dissection method, allowing for a more nuanced interpretation of market dynamics.

Empirically, our results painted a multifaceted picture. Using databases of Bloomberg, we collected extensive data on various indices including the S&P 500, DAX 30, and FTSE 100 index. Through our in-depth analysis, we observed distinct patterns and variations in volatility. Models like GARCH, GJR-GARCH, HAR, and HAR-a, along with their extensions, provided insights into the forecast performance of different methods. Specifically, while models like GARCH and GJR-GARCH solely relied on daily prices, the inclusion of intraday data in HAR and HAR-a led to a closer match with realized volatility, thereby affirming the power of high-frequency data in volatility prediction. Iterated cluster partition method applied to HAR and HAR-a can enhance the volatility forecasting even better, especially the high-volatility period like 2019-2020.

The value-at-risk (VaR) analysis shed light on the risk estimation capabilities of various models. While models like HAR and HAR-a show excellent volatility forecasts, they tended to overestimate risk. Iterated cluster partition method applied on GARCH and GJR show best VaR, reflecting the nuances and intricacies of volatility forecasting. Our out-of-sample forecasts further emphasize the importance of multiple window sizes and robustness checks, leading to a more comprehensive understanding. While certain models demonstrated superiority in specific contexts, the essence is that a one-size-fits-all approach may not be ideal in predicting market volatility.

In conclusion, our exploration of cluster volatility offers a fresh take on understanding market volatility. While our empirical findings underscore the importance of selecting the right model for the right context, our theoretical underpinnings promise a new horizon for future research in financial econometrics. The convergence analysis of the iterated cluster partition model and its applicability to multivariate time series will be our further research targets in the future. In terms of application, not only the volatility series of index or stock returns, but also the volatility of other assets or the implied volatility under risk-neutral measure can be potential application scenarios of cluster

partition volatility.

References

Andersen, T.G., Bollerslev, T., (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39, 885-905.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., (2006). Volatility and correlation forecasting. Handbook of economic forecasting, 1, 777-878.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579-625.

Andreou, E., Ghysels, E., (2002). Detecting multiple breaks in financial market volatility dynamics. Journal of Applied Econometrics, 17(5), 579-600.

Andreou, E., Ghysels, E. (2009). Structural breaks in financial time series. Handbook of financial time series, 839-870.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., (2006). The cross-section of volatility and expected returns. Journal of Finance, 61, 259-299.

Angelini, G., Bacchiocchi, E., Caggiano, G., Fanelli, L., (2019). Uncertainty across volatility regimes. Journal of Applied Econometrics, 34(3), 437-455.

Bai, J., Perron, P., (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics, 18(1), 1-22.

Barndorff-Nielson, O.E., Shephard, N., (2002). Estimating quadratic variation using realised variance. Journal of Applied Econometrics, 17(5), 457-477.

Berkowitz, J., Christoffersen, P., Pelletier, D., (2011). Evaluating value-at-risk models with desk-level data. Management Science, 57(12), 2213-2227.

Bollerslev, T., (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327.

Brown, M.B., Forsythe, A.B., (1974). Robust tests for the equality of variances. Journal of the American Statistical Association, 69(346), 364-367.

Calvet, L., Fisher, A., (2004). How to forecast long-run volatility: Regime-switching and the estimation of multifractal processes. Journal of Financial Econometrics, 2, 49-83.

Campbell, J.Y., Giglio, S., Polk, C., (2018). An intertemporal CAPM with stochastic volatility. Journal of Financial Economics, 128(2), 207-233.

Carrasco, M., Hu, L. Ploberger, W., (2014). Optimal Test for Markov Switching. Econometrica, 82, 765-784.

Chen, Y., Wang, Z.C., Zhang, Z.J., (2019). Mark to market value at risk. Journal of Econometrics, 208(1), 299-321.

Cheng, Y., (1995). Mean shift, mode seeking, and clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8), 790-799.

Chkili, W., Hammoudeh, S., Nguyen, D.K., (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Economics, 41, 1-18.

Chuffart, T., (2017). An implementation of markov regime switching garch models in matlab. Available at SSRN 2892688.

Chung, K. H., Wang, J., Wu, C., (2019). Volatility and the cross-section of corporate bond returns. Journal of Financial Economics, 133(2), 397-417.

Coleman, G.B., Andrews, H.C., (1979). Image segmentation by clustering. Proceedings of the IEEE, 67(5), 773-785.

Conrad, C., Kleen, O., (2020). Two are better than one: Volatility forecasting using multiplicative component garch-midas models. Journal of Applied Econometrics, 35, 19-45.

Corsi, F., (2008). A simple approximate long-Memory model of realized volatility. Journal of Financial Econometrics, 7(2):174-196.

Diebold, F.X., Mariano, R.S., (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13, 134.

Ederington, L.H., Lee, J.H., (1993). How markets process information: news releases and volatility source. Journal of Finance, 48(4), 1161-1191.

Engle, R.F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987-999.

Engle, R.F., Ng, V.K., (1993). Measuring and testing the impact of news on volatility. Journal of Finance, 48(5), 1749-1778.

Engle, R.F., Manganelli, S., (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. Journal of Business & Economic Statistics, 22(4), 367-381.

Engle, R.F., Siriwardane, E.N., (2018). Structural GARCH: The volatility-leverage connection. Review of Financial Studies, 31(2), 449-492.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd, 96(34), 226-231.

Garcia, R., (1998). Asymptotic null distribution of the likelihood ratio test in markov switching models. International Economic Review, 39, 763-788.

Glosten, L.R., Jagannathan, R., Runkle, D.E., (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48(5), 1779-1801..

Haas, M., Mittnik, S., Paolella, M.S., (2004). A new approach to Markov-switching GARCH models. Journal of Financial Econometrics, 2(4), 493-530.

Hamilton, J.D., (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 2(4), 357-384.

Hamilton, J.D., Susmel, R., (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime. Journal of Econometrics, 64(1-2), 307-333.

Hamilton, J.D., (2010). Regime-Switching models. Macroeconometrics and time series analysis. Palgrave Macmillan, London, 202-209.

Hansen, B. E., (1992). The Likelihood Ratio Test under Non-Standard Conditions. Journal of Applied Econometrics, 7(S1), S61-S82.

Hansen, P.R., Huang, Z., Shek, H.H., (2012). Realized garch: a joint model for returns and realized measures of volatility. Journal of Applied Econometrics, 27(6), 877-906.

Jain, A.K., Murty, M.N., Flynn, P.J., (1999). Data clustering: a review. ACM Computing Surveys (CSUR), 31(3), 264-323.

Klaassen, F., (2002). Improving GARCH volatility forecasts with regime-switching GARCH. Empirical Economics, 27, 363-394.

Koop, G. and Potter, S. M., (2009). Prior elicitation in multiple change-point models. International Economic Review, 50(3), 751-772.

Kupiec, P.H., (1995). Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives, 3(2), 73-84.

Lee, T., Kim, M., Baek, C., (2015). Tests for volatility shifts in GARCH against long-range dependence. Journal of Time Series Analysis, 36(2), 127-153.

Liu, C., Maheu, J.M., (2008). Are there structural breaks in realized volatility? Journal of Financial Econometrics, 6(3), 326-360.

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

Orhan, M., Köksal, B., (2012). A comparison of GARCH models for VaR estimation. Expert Systems with Applications, 39(3), 3582-3592.

Patton, A.J., (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160, 246-256.

Pelletier, D., (2006). Regime switching for dynamic correlations. Journal of Econometrics, 131(1-2), 445-473.

Riordan, R., Storkenmaier, A., (2012). Latency, liquidity and price discovery. Journal of Financial Markets, 15(4), 416-437.

Schmitt, N., Frank, W., (2017). Herding behaviour and volatility clustering in financial Markets. Quantitative Finance, 17(8), 1187-1203.

Shi, B, Bai, X, Yao, C., (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11), 2298-2304.

Sims, C. and Zha, T., (2006). Were There Regime Switches in U.S. Monetary Policy? American Economic Review, 96(1), 54-81.

Smetanina, E., Wu, W.B., (2021). Asymptotic theory for qmle for the real-time garch(1,1) model. Journal of Time Series Analysis, 42, 752-776.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., (2010). Detecting trend and seasonal changes in satellite images time series. Remote Sensing of Environment, 114(1), 106-115.

Xu, K.L., (2013). Powerful tests for structural changes in volatility. Journal of Econometrics, 173, 126-142.

Zhang, G.P., Baek, C., (2019). Neural networks for classification: a survey. IEEE Transactions on Systems, Man, and Cybernetics, 30(4), 451-462.

Author Statement

1. Author 1: Data curation, Software, Formal Analysis, Methodology, Visualization, Writing - Original Draft, Funding acquisition
2. Author 2: Supervision, Conceptualization, Methodology, Writing - Review & Editing.
3. Author 3: Validation, Investigation.
4. Author 4: Project administration, Writing - Review & editing.

Volatility Forecasts by Clustering: Applications for VaR Estimation

Zijin Wang
School of Economic Mathematics
Southwestern University of Finance and Economics
Chengdu, 611130, P.R. China
Tel: +86-15884554480
E-mail: 1190202Z1006@smail.swufe.edu.cn

Peimin Chen*
Business Analytics Programme
Division of Business and Management
BNU-HKBU United International College
Zhuhai, Guangdong, 519087, China.
Tel: +86-18328516475
E-mail: peiminchen@uic.edu.cn

Peng Liu
S.C. Johnson College of Business
Cornell University
Ithaca, NY, 14853, USA
Tel: 510-277-2874
Fax: 607-254-2960
E-mail: peng.liu@cornell.edu

and

Chunchi Wu*
School of Management
State University of New York
Buffalo, NY, 14260, USA
Tel: 716-645-0448
Fax: 716-645-3823
E-mail: chunchiw@buffalo.edu

Current version: March 19, 2024.

*Corresponding author.