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*
Shanghai Business School
Shanghai, 200235, P.R. China
Tel: +86-18328516475
E-mail: pmchen@sbs.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: June 5, 2021.

*Corresponding author.

Volatility Forecasts by Clustering: Applications for VaR Estimation

Abstract

It is well known that volatility is time-varying and clustered. However, little research has explored the information content of volatility clustering and its implications for investment decisions. This information is particularly important in turbulent periods such as Financial Crisis. We present a volatility cluster partition model to forecast volatility and apply it to risk management. We find that our model substantially outperforms the GARCH model and improves financial risk management using the value-at-risk metric.
Keywords: Volatility forecasts; Fisher's optimal dissection; value-at-risk;

1. Introduction

Return volatility is an important risk variable in asset pricing, portfolio selection and risk management (see Angelini et al. (2009); Schmitt and Frank (2017); Engle and Siriwardane (2018); Chen et al. (2019)). To capture instantaneous volatility dynamics and clustering, the classic Autoregressive Conditional Heteroskedasticity (ARCH) model and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model proposed by Engle (1982) and Bollerslev (1986) are popular for its simple form and easy to explain. Many extension works such as Nelson (1991), Engle and Ng (1993)) examined properties of assymmetry, long persistency and so on. However, for most dynamic volatility models, it is hard to estimate their parameters jointly using the real data, because many factors are involved and there are substantial noises. these GARCH models also have shortcomes. Simple forms of ARCH and GARCH models are out-of-style and has been beaten by more elaborate GARCH-family models. As these models envolve more complex, it is harder to estimate their parameters using real data. Especially for update data, estimation is very time consuming and inconvenient. For application purposes, it is therefore desirable to have a simple volatility measure, which can impound current market information promptly. In this paper, we propose a volatility cluster partition model based on clustering analysis method to forecast return volatility. This model

captures market information in real time and is easy to implement. Applying our model forecasts to financial risk management using the value-at-risk (VaR) metric, we find that it outperforms conventional methods by a substantial margin.

Stock prices can change abruptly and it is important to accommodate this data feature in volatility forecasts. As an example, since its introduction, the circuit breaker has been triggered only once on October 27, 1997, when the Dow Jones industrial index fell 7.18%, before 2020. However, from March 9 to March 18, 2020, the circuit breaker was triggered four times. Particularly, on March 16, the S&P 500 index triggered a circuit breaker at the market open for the first time in history, and the index experienced the largest one-day decline in nearly 33 years with a drop of 11.98%. Over this period, the market volatility is extremely high. Ignoring the structure shift in time series can lead to inefficient volatility forecasts.

Volatility exhibits strong clustering, i.e., volatility remains stable over a period of time but can suddenly switch to another state. When volatility is stable, time-variation is a lesser concern. Thus, we can use the average value of volatility during a stable period, and determine the optimal time points for cluster partitions. Likewise, we can update the value of volatility in another stable period in a similar way. Schmitt and Frank (2017) assume herding behaiour of speculators leads to volatility clustering. In calm time, investors make decisions independently. During turbulent period, market participants are sensitive to others' behaviour and they are located on either the buy side or the sell side. With unbalenced structure, stock prices adjust dramaticly and volatility remain high.

The regime of market either calm or turbulent is persistent, however, the status can change if there appear significant signals for investers, such as Financial Crisis, wars or government policies. This can be driven by a regime-switching(RS) process. For RS models, a Markove process with state-dependent transition probabilities governs the switching between regimes. The maximum likelihood method is employed for the statistical inference on the optimal regime. Hamilton (1989) introduces this model to describe the U.S. business cycle, which is characterized by periodic shifts from recessions to expansions and vice versa. Klaassen (2002) proposes RS-GARCH models, in which GARCH is used in each regime to describe the volatility process. Pelletier (2006) proposes an alternative volatility model for multiple time series with the regime switching

dynamic correlation. The covariance is a constant within a regime but changes across regimes.

Signals of RS models can be phrased as structural change points, which has been traditionaly detected by hypothesis tests (see, for example, Andreou and Ghysels (2002) and Lee et al. (2015)). Brown and Forsythe (1974) modify the Levene test on the equality of ex-ante and ex-post variance changes and this test is widely used in studies of stock prices (see Riordan (2012)) and futures markets (see Ederington and Lee (1993)).

Clustering annlysis has been popular employed these years, such as k-mean (see MacQueen (1967)), mean-shift clustering algorithm (see Cheng (1995)), and density-based spatial clustering with noise (DBSCAN, see Ester et al. (1996)), having been developed and extended in many applied fields from statistics (see Jain (1999)) to image segmentation (see Coleman (1979)). More complicated methods with parameters, including support-vector machine and neural networks, are developed in artificial intelligence (see, for example, Zhang (2000); Shi (2016)). It has been shown that these methods are efficient in processing the data with no time order. However, in most financial models, the data involved follows a time series process and it is therefore necessary to find an alternative method to cluster the series. The Fisher optimal dissection method is a suitable algorithm for financial modeling as it can cluster dynamic volatility by minimizing a loss function defined by the similarity of samples. The basic idea of such a method is widely employed in economics (see Bai and Perron (2003)) and image treatment (see Verbesselt et al. (2010)).

Thus, we can find the time points for volatility clustering and use representative values of volatilities during a stable period. For a new update data, we can easily examine whether it is balongs to the latest cluster. If it is, then the representive value of this cluster can be used to predict the future volatility. If not, it is a signal of regime switch. .

Our model is differentiated from the conventional models for structural changes in two major aspects. First, our model performs optimal volatility partitions and generates timely classification by incorporating the most updated information. To identify the optimal classification, two or more samples are sufficient for implementing our method. Thus, our model can capture market information much faster. This procedure contrasts with the tests commonly used in prior researches that rely on the statistical inference on

the parameters of the model. In these tests, the occurrence of the structural change in parameters only reveals after enough samples are produced. As such, it cannot capture instantaneous fluctuations in volatility. To catch market information flows immediately, a more efficient method is needed. Second, the algorithm we develop is convenient and general, and there is no need to derive the inference for test statistics. It is not easy to draw an efficient statistical inference for the structural change or regime-switching model of the implied volatility calculated from the option pricing formula by the iteration method. Our algorithm incorporates the newly calculated implied volatility into the model to quickly infer structural changes.

The GARCH model has been widely used for volatility forecasts by academicians and practitioners. Orhan and Köksal (2012) compare various GARCH models for quantifying the VaR in times of stress. VaR measures the maximum likely loss of an investment portfolio within a certain time period at a confidence level. Other methods have also been used to calculate the VaR, i.e., variance-covariance, Monte Carlo, and nonparametric methods. In this paper, we choose GARCH as a benchmark to evaluate the performance of our model in VaR applications.

This paper makes several contributions to the current literature. First, we propose a new model to cluster dynamic volatility using Fisher's optimal dissection method. A unique feature of this model is that the simulated volatility is a constant in each cluster section but promptly responds to shocks at the cluster partition point. Second, unlike previous models, estimating the parameters of our model is relatively straightforward. Third, our model outperforms the traditional GARCH model in volatility forecasts. Using the volatility simulated by our model, we find that it significantly improves risk management using the VaR metric. The clustering volatility estimated by our model tracks real volatility more closely than the standard GARCH model.

The remainder of this paper is organized as follows. In Section 2, we introduce the models of GARCH dynamic volatility.In Section 3, we present the cluster model based on Fisher's optimal dissection method. In Section 4, we discuss the applications of our volatility measures to VaR and option price predictions and show that our method enhances the forecast of the GARCH-MIDAS model. In Sections 5 to 7, we provide examples of VaR simulation, option pricing prediction and improved GARCH-MIDAS forecasts. Finally, in Section 8, we summarize our main findings and conclude the paper.

2. Volatility cluster model

2.1. GARCH volatility model

In this section, we simply start from a typical ARCH/GARCH model taking the form:

$$r_t = \sigma_t \varepsilon_t,\ \varepsilon_t \sim N(0,1) \tag{1}$$

where r_t is the return at time t, ε_t a shock with mean zero and variance one and σ_t is the conditional volatility of returns conditioned on past information $\mathcal{F}_{t-1}$ as:

$$\begin{aligned} \sigma_t &= \mathbb{E}[r_t^2|\mathcal{F}_{t-1}] \\ &= \mathbb{E}[r_t^2|r_{t-1},\ldots,r_1,\sigma_{t-1},\ldots,\sigma_1]. \end{aligned} \tag{2}$$

This conditional expectation is typically parametrized as a recursive process, in a standard GARCH model

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

where ω is parameter, α is ARCH parameter and β is GARCH parameter. Later we add more elaborated recursive structures, see Nelson (1991) and Glosten, Jagannathan and Runkle (1993).

For GJR-GARCH model, the conditional volatility is

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

where $\mathbb{1}_{r_{t-1}<0}$ is an indicator function, which equals to 1 when $r_{t-1} < 0$ and equals 0 otherwise.

The problem of GARCH models is that these models can not include latest volatility timely. For example, when a new data comes which is a big drop or big increase, the model will just take it into estimation for the whole data sample. This new data cannot change the parameter estimated significantly and the model cannot detect what kind of change it is, whether it is an outlier of innovition or a change of volatility.

An alternative way is to model directly volatility using realized volatility.

2.2. HAR-GARCH

RV is constructed from the sum of intraday squared returns. Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002) discuss the precision of RV.

Market microstructure dynamics contaminate the price process with noise. The noise can be time dependent and may be correlated with the efficient price (Hansen and Lunde, 2006). RV can be a biased and inconsistent estimator. For instance, see Bandi and Russell (2006), Hansen and Lunde (2006), Oomen (2005), and Zhang et al. (2005) for more details on the effects of market microstructure noise on volatility estimation. To reduce the effect of market microstructure noise, Liu and Maheu (2008) employ a kernel-based estimator that utilizes autocovariances of intraday returns. Specifically, Liu and Maheu (2008) follow Hansen and Lunde (2006) to provide a bias correction to realized volatility as follows

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

where $r_{t,i}$ is the ith logarithmic return during day t, M is the number of returns in each day, and $q = 1$ in this calculation.

Consider the heterogeneous autoregressive model (HAR) of realized volatility by Corsi (2008) and Liu and Maheu (2008), which can capture many of the features of volatility including long memory. The logrithmic of the HAR is defined as:

$$v_t = b_0 + b_1 v_{t-1} + b_2 v_{t-5,t-1} + b_3 v_{t-22,t-1} + \epsilon_t,\ \epsilon_t \sim NID(0, \sigma^2), \tag{5}$$

where $v_t = \log(RV_t)$ and

$$v_{t-5,t-1} = \frac{\log(RV_{t-1}) + \log(RV_{t-2}) + \cdots + \log(RV_{t-5})}{5}, \tag{6}$$

$$v_{t-22,t-1} = \frac{\log(RV_{t-1}) + \log(RV_{t-2}) + \cdots + \log(RV_{t-22})}{22}. \tag{7}$$

This model postulates three factors that affect volatility: daily log-volatility v_{t-1}, weekly log-volatility $v_{t-5,t-1}$, and monthly log-volatility $v_{t-22,t-1}$.

Including asymmetry and junps, model could be

$$\begin{aligned} v_t =& b_0 + b_1 v_{t-1} + b_2 v_{t-5,t-1} + b_3 v_{t-22,t-1} + b_J J_{t-1} + b_{A1} \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \\ &+ b_{A2} \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \mathbb{1}_{r_{t-1}<0} + \epsilon_t, \end{aligned} \tag{8}$$

where J_{t-1} is a jump component defined as Liu and Maheu (2008)

$$J_t = \begin{cases} \log(RV_t - RBP_t + 1) & when\, RV_t - RBP_t > 0 \\ 0 & otherwise \end{cases} \tag{9}$$

where $RBP_t = \frac{\pi}{2} \sum_{i=1}^{M-1} |r_{t,i}||r_{t,i+1}|$ is realized bipower variation.

2.3. Regime Switching GARCH

Following the basic framework as

$$r_t = \sigma_{s_t} \varepsilon_t \tag{10}$$

where s_t is a random variable of state at date t. The simplest specification is a two-state Markov chain, which in general can extent to N states as

$$Pr(s_t = j | s_{t-1} = i, s_{t-2} = k, \cdots, r_{t-1}, r_{t-2}, \cdots) = Pr(s_t = j | s_{t-1} = i) = p_{ij}, \tag{11}$$

where $i, j \in 1, \cdots, N$. Most applications assume only $N = 2$ or 3 different regimes though there are considerable models with a much larger number of regimes, either by tightly parameterizing the relation between regimes see Calvet and Fisher (2004), or with prior Baysian information Sims and Zha (2006).

Haas et al. (2004) poses the model where there are N separate GARCH processes whose conditional volatility σ_{it} all exist as latent variables at date t

$$\sigma_{it}^2 = \omega + \alpha_i r_{t-1}^2 + \beta_i \sigma_{i,t-1}^2 \tag{12}$$

where σ_{it}^2 is a function solely on V_{t-1} rather than the states $s_{t-1}, s_{t-2}, \cdots, s_1$.

3. Cluster Partition

GARCH family models usually contains infinite order information about conditional variances and errors, which may be irrelavant and even disturbing, especially when volatility changes dramatically. For example, by the sudden epidemic impact, the dramatic fluctuation of U.S. stock market volatility in March 2020 and 4 circuit breakers left investors with huge losses. If GARCH family models are used to forecast volatility, it will be underestimated due to the ineffectiveness of previous information. If forecast is based only on more recent data, precision will improve. Our model is to consider this point, through clustering we can find the most relevant time interval with the current time, and then use its representative cluster value to predict the volatility in short future.

$V = (\sigma_1, \ldots, \sigma_T)$ are conditional variance estimated by a certain GARCH model, for example GJR-GARCH considering asymmetry[1] (see Glosten, Jagannathan and Runkle (1993)), V is divided into N clusters $C_1, \ldots, C_N$, and the representative value $\bar{\sigma}_N$ of the last cluster can be used as forecast σ_{T+1}. This forecast has good properties of both original GJR-GARCH model and cluster partition model.

3.1. Single structural change statistic

In time-series, it is important for researchers and financial analysts to discern a sequence transferring from one cluster to another as time goes by. This concern leads to two problems, known and unknown structural breaks. Traditional structural break problems have been phrased as hypothesis tests. The null is set up to describe structural stability, the alternative contains one or multiple structural break(s). Classical Chow Test (Chow (1960)) constructed F statistic by linear regression before and after a structural break. Another popular way is to introduce dummy variable, and test all dummy variables and the interaction coefficient between dummy variables and independent variables.

For unknown structural breaks, select a subinterval and test every point in the subinterval to derive the corresponding F statistic. Take the largest F statistic of them and then select the corresponding point as structural break. This F statistic is also known as Quandt Likelihood Ratio(QLR).

[1] Other GARCH family models can also be implemented.

In general, the test statistics may be viewed as two-sample tests adjusted for the unknown break location. Often asymptotic relationships are derived to obtain critical values, see (Bai and Perron (1998)).

To identify this pattern, we construct a statistic for cluster partitions. A partition point is selected with the minimum information loss, such that individual observations within the same cluster are alike while the difference between clusters is apparent.

Let I be a partition point. Then, divide the original data into two clusters if we have[2]

$$\sum_{i=1}^{n}(\sigma_i - \bar{\sigma}_i)^2 > \sum_{i_1=1}^{I-1}(\sigma_{i_1} - \bar{\sigma}_{i_1})^2 + \sum_{i_2=I}^{n}(\sigma_{i_2} - \bar{\sigma}_{i_2})^2, \tag{13}$$

where $\bar{\sigma}_i$, $\bar{\sigma}_{i_1}$ and $\bar{\sigma}_{i_2}$ are the corresponding gathering centers of $\{\sigma_i\}_{i=1}^{n}$, $\{\sigma_{i_1}\}_{i_1=1}^{I-1}$ and $\{\sigma_{i_2}\}_{i_2=I}^{n}$, respectively, $I = 2, \cdots, n$. Generally, the arithmetic or geometric mean values of the corresponding samples including the given element are used. To have an algorithm of fast speed and a simple form of model, we employ arithmetic mean values as gathering centers. The greater the difference between the left side and right side, the more likely for two different clusters. Since the sum of the left side of equation (13) is fixed, our aim is to minimize the right side. We construct the cluster statistic as

$$\rho_\sigma(I) = \frac{1}{T}\sum_{i_1=1}^{I-1}(\sigma_{i_1} - \bar{\sigma}_{i_1})^2 + \frac{1}{T}\sum_{i_2=I}^{n}(\sigma_{i_2} - \bar{\sigma}_{i_2})^2. \tag{14}$$

By minimizing $\rho_\sigma(I)$, at least one partition point I can be identified. There may be more than one detected partition points with same statistic value $\rho_\sigma(I_1)$ and $\rho_\sigma(I_2)$.

To illustrate the cluster algorithm, as an example, we select the volatility data with 252 daily observations in 2018. Panel A of Figure 1 shows the cluster statistic for every possible partition point. Cluster statistics peak at $(197, 0.0012)$ which suggests that the optimal partition point is the 197-th in the sample, or October 12, 2018. This result is consistent with the pattern in Panel B, where volatility jumps to a high value after October 12. In contrast, volatilities before October 12 mostly stay in relatively low level. Though two peaks, located at 23-th (February 5) and 79-th (April 26) in Panel A,

[2]In fact, we can construct a statistic inference by Chi-square test for the series $\{v_i\}_{i=1}^{n}$ to search its partition point for two clusters. But here we adopt the simple form as the inequality (13) to be consistent with the latter Fisher's optimal dissection method.

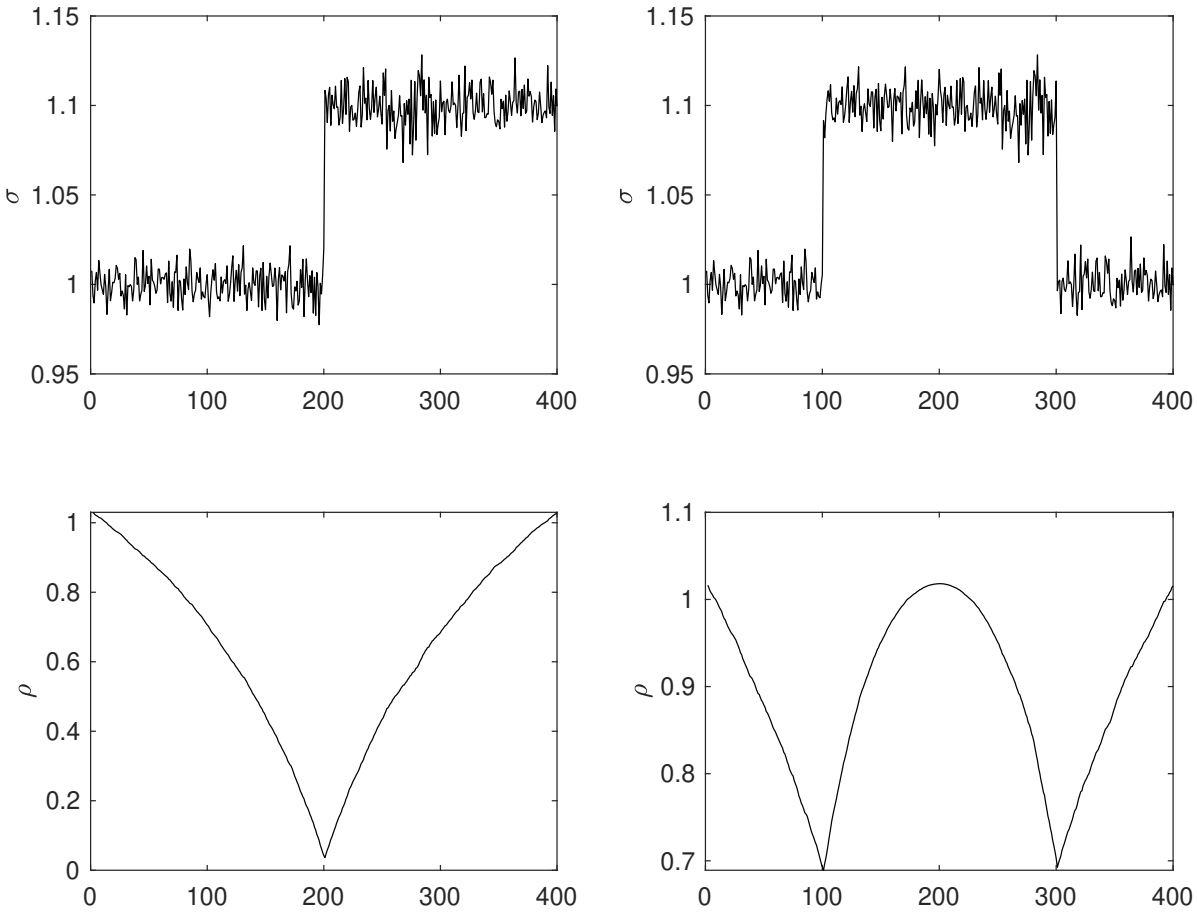


Figure 1: Single Structural Change Statistic

exist and they represent two sub-optimal partition points spliting the sample into two distinguishing parts, they are not as good as the optimal partition point corresponding to 197-th (October 12).

3.2. General cluster partitions of dynamic volatility

Similarly, we can construct the cluster statistic for more than two clusters using a forward algorithm. For volatility series V, a classifier $\pi : \mathbb{R}^T \mapsto \mathbb{R}^N$ denoted as $\pi(V, N)$, divides V into N clusters $(C_1, \ldots, C_N)$ by N split points $(I_1, I_2, \ldots, I_N)$, where $1 = I_1 < I_2 < \cdots < I_N < T$ (define $I_{N+1} = T + 1$ for convenience). Among these clusters, the index of the first point in each cluster C_k is I_k, for $k = 1, \ldots, N$. In fact, $\pi(V, N)$ consists of N split points with only the first point I_1 always fixed, suggesting $\pi = (I_1, I_2, \ldots, I_N) \in \mathbb{N}^+_{N-1}$.

Table 1: Dive V into N clusters

Cluster	First point index	Components
C_1	I_1	$\sigma_1, \cdots, \sigma_{I_2-1}$
C_2	I_2	$\sigma_{I_2}, \cdots, \sigma_{I_3-1}$
$\vdots$	$\vdots$	$\vdots$
C_N	I_N	$\sigma_{I_N}, \cdots, \sigma_n$

Note: Given cluster number N.

Define loss function of classifier π as

$$L(\pi) = \frac{1}{T} \sum_{k=1}^{N} D(C_k), \tag{15}$$

where $D(C_k) = \sum_{i \in C_k} d(\sigma_i, \bar{\sigma}_{C_k})$ the diameter of cluster k with $\bar{\sigma}_{C_k}$ the central of cluster C_k and $d(\cdot)$ the distance function, here we take the Euclidean distance[3]. The optimal classifier is the one with lowest loss function value

$$\pi^{CP}(V, N) = \arg \min_{\pi \in \mathbb{N}_{N-1}^{+}} L[\pi(V, N)]. \tag{16}$$

The optimal clusters are determined by the optimal classifier. The problem of finding multiple optimal partition points in a fixed interval is essentially the extension of the binary classification problem we introduce in the last subsection. Assuming that the first part of the interval has been optimal already, it is a binary classification problem to find a single optimal partition point for the remaining part. Then the interval can be divided into a subinterval which has achieved optimal segmentation and a subinterval with one single partition point to be determined by binary classification. Therefore, the original optimization can be transformed into a recursive binary classification problem. A forward iteration algorithm provides the optimal clusters and the optimal classifier $\pi^{CP}(V, N)$ (see Appendix A for more details). In this sense, the corresponding optimal split points are $(I_1^{CP}, \ldots, I_N^{CP})$. The classifier π^{CP} consists of these N split points, which are only dependent on conditional volatility V and cluster number N. This clas-

[3]The distance measure can be generated by L_p-norms. The Euclidean distance is L_2-norm, taken as $d(\sigma_i, \bar{\sigma}_{C_k}) = \|\sigma_i - \bar{\sigma}_{C_k}\|_2$.

sifier is optimal because conditional volatility is homogeneous in each cluster and shifts apparently across clusters, which serves as a good description of volatility cluster.

3.3. Cluster Number

As cluster number N affects the optimal classifier, an important task is to determine a proper cluster number.

In traditional models, a natural way is to test the null hypothesis that there are N clusters against the alternative of $N+1$[4]. This test, as Hamilton (2010) says, fails to satisfy the usual regularity conditions because under the null hypothesis, some of the parameters of the model would be unidentified[5]. To interpret a likelihood ratio statistic one instead needs to appeal to the methods of Hansen (1992) or Garcia (1998). Alternative tests are not based on likelihood ratio statistic[6]. Other alternatives are to use Bayesian methods to calculate the value of N implying the largest value for the marginal likelihood (see Liu and Maheu (2008)), the Bayesian Factors (see Koop and Potter (2009)), or compare models on the basis of their ability to forecast (see Hamilton and Susmel (1994)). However, these methods can only detect regimes no more than 10. In pratice, the volatility evolves much more dramatically especially in crisis. We need a more effective methods to detect the cluster number.

There are two methods to determine this value in the literature: the graphical method and the slope method. For the graphical method, the loss function $L\left[\pi\left(V,N\right)\right]$ is figured as a function of cluster number N. The N number with the largest turning amplitude is chosen as the optimal cluster number. For the slope method, a non-negative slope is constructed and its inflection point is identified as the optimal cluster number.

The graphical method is simple and intuitive, suitable for a small sample cluster. However, when the number of samples grows, there are many potential curve turning points, and it becomes more difficult to judge the optimal cluster number by observation.

The slope statistic to determine the optimal cluster number is given by

$$S(N)=\frac{L\left[\pi\left(V,N\right)\right]}{L\left[\pi\left(V,N+1\right)\right]}.$$

[4]See some references here

[5]For example, if there is really only one cluster for the whole sample, the maximum likelihood estimate does not converge to a well-defined population magnitude, meaning that the likelihood ratio test does not have the usual χ^2 distribution.

[6]See Carrasco et al. (2004)

As $S(N)$ is closed to 1, the cluster number approaches optimality. The slope statistic less than 1 suggests no need for a bigger N. Nevertheless, slope statistic may approach but never get less than 1, and too many pieces of clusters may cause an over-fitting problem. Therefore, it is important to select an optimal cluster number exactly.

Concerned with the information criterion for model selection, we introduce an information-based statistic, which provides insight into its relationship to the optimized $L\left[\pi\left(V,N\right)\right]$ and the penalty for complexity, defined by

$$\psi(N)=\log(L\left[\pi\left(V,N\right)\right])+\frac{N\log(T)}{T}, \tag{17}$$

where $\frac{N\log(T)}{T}$ serves as a penalty. The minimum information-based statistic leads to the optimal cluster number.

3.4. Cluster partition volatility

Based on cluster partition method, V is divided into N clusters $C_1, C_2, \cdots, C_N$, of which volatility difference is small within each cluster and large cross clusters. This is very important for volatility clustering. In each cluster, conditional homogenaity is easily satisfied. Cluster partition volatility $V^{CP}=(\sigma_1^{CP},\sigma_2^{CP},\ldots,\sigma_T^{CP})$ can be expressed as form of clusters $(C_1, C_2, \ldots, C_N)$. For each cluster k, a typical value $\overline{\sigma}_k$ could be a representative cluster value,

$$\sigma_t^{CP}=\mathbb{1}_{\{\sigma_t\in C_k\}}\overline{\sigma}_k, \tag{18}$$

where $k=1,\cdots,N$, $\mathbb{1}_{\{\cdot\}}$ is indicator function and σ_t^{CP} is the conditional volatility by cluster partition method. In the following, we set $\overline{\sigma}_k$ to be a constant as the arithmetic average of corresponding cluster. In fact, there are other forms of constant representative volatility, and we adopt the simplest one as above for convenience. Later we will see different forms of average make little difference. From the homogeneous property, this partitional cluster volatility can be applied to forecasting volatility.

4. Cluster volatility in financial models

In this section, we present an applications of cluster volatility, the value-at-risk (VaR) used in risk management.

4.1. VaR and estimation

VaR is important in financial risk management mainly due to its simple form and easily interpretable feature. Based on its definition, VaR is the highest possible loss of a financial asset over a given time interval, with a given level of confidence. It gained a higher profile in 1994 when J.P. Morgan published its RiskMetrics system. The Basel Committee on Banking Supervision proposed in 1996 that internal VaR models may be used in the determination of the capital requirements that banks must fulfill to back their trading activities (Chen et al. (2019)).

For returns $r_{t+1}, \cdots, r_{t+l}$ in l consecutive trading days, denote $S_{t,l} = \sum_{i=t+1}^{t+l} r_i$ as the cumulative return at the end of the day $t+l$, where l has often been chosen as 1, 10, 20, 60, correspding to 1 day, 2 weeks, 1 month and 3 months respectively. Let $P(S_{t,l} \leq x)$ be the probability of l-day cumulative return no more than x and $F_{S_{t,l}}(x) = P(S_{t,l} \leq x)$ is the distribution function of $S_{t,l}$. For a significant level α $(0 < \alpha < 1)$, a l-day VaR is defined by

$$\mathrm{VaR}_{t,l}(\alpha) = -\sup\left\{x : F_{S_{t,l}}(x) \leq \alpha\right\}. \tag{19}$$

There exist three main methods to in estimating VaR, i.e., historical simulation approach, mean variance approach and model-based analytical method (Chen et al. (2019)).

For historical simulation, we suppose that returns will repeat in the future. Denote $Y_{t,l} = \sum_{i=t-l}^{t-1} r_i$ as the cumulative return of l lags at the end of day t. The empirical cumulative distribution function (cdf) of $F_{S_{t,l}}(y)$ is

$$\widehat{F}_{S_{t,l}}(y) = \frac{1}{n} \sum_{\tau=t-n}^{t-1} \mathbb{1}\left(Y_{\tau,l} < y\right) \tag{20}$$

and the estimation of VaR is

$$\widehat{VaR}_{t,l}(\alpha) = -\sup\left\{x : \widehat{F}_{S_{t,l}}(x) \leq \alpha\right\}. \tag{21}$$

For mean variance approach, the VaR at time t under normal distribution can be written as

$$\mathrm{VaR}_{t,l}(\alpha) = \hat{\mu}_t + z_\alpha \hat{\sigma}_t \tag{22}$$

where z_α designates the left quantile at α of the normal distribution, $\hat{\mu}_t$ and $\hat{\sigma}_t$ refer to

the estimated daily conditional mean and volatility obtained from based model. When assuming a Student-t distribution, the VaRs are given by

$$\mathrm{VaR}_{t,l}(\alpha) = \hat{\mu}_t + st_{\alpha,\nu}\hat{\sigma}_t \tag{23}$$

where $st_{\alpha,\nu}$ is the left quantile at α of the student-t distribution with ν degrees of freedom.

For each return series of give time period, we compute the VaR at several pre-specified significance level of α from 0.5% to 5%. We then examine the performance of the considered models by calculating the empirical failure rate of the return distributions. Here the failue rate is defined as the number of times the return series. If the failure rate is equal to the pre-specified VaR level, we can then conclude that the associated VaR model is correctly specified. This hypothesis is explicitly tested by the Kupiec Likelihood Ratio (LR) test (Kupiec (1995)). The statistic of the Kupiec LR test is given by

$$LR = -2\ln\left[\alpha^{T-N}(1-\alpha)^{N}\right] + 2\ln\left[(1-f)^{T-N} f^{T}\right] \tag{24}$$

where f denotes the empirical failure rate calculated as the ratio of the number of return observations exceeding the estimated VaR value (N) to the sample size (T). The Kupiec LR statistic is asyptotically chi-squared distributed with one level of freedom under the null hypothesis that the failure rate f equals the pre-specified confidence level α.

4.2. Data selection

We collect daily closed prices P_t of the S&P 500 index from January 3, 1989 to December 31, 2018. Returns are measured by logarithmic price changes, $r_t = \ln(P_t/P_{t-1})$. To compare the results from different methods, we focus on a single financial asset with daily data, and the confidence levels are set at 99.5%, 99%, 97.5%, 95% 92.5% and 90%, respectively. To evaluate the accuracy of VaR calculation, two samples with short and long time horizons are used. For the short-horizon sample, the VaR values are calculated over a 2-year period with 501 trading days, starting from January 3, 2017 to December 31, 2018. For the long-horizon sample, we simulate VaR values over a 10-year period with 2,515 trading days, spanning from January 3, 2009 to December 31, 2018. The number of failure events is compared at the same confidence level to judge the merits of different models.

4.3. Cluster volatility and VaR

To describe the cluster phenomenon of volatility, we first consider the GARCH model before using the volatility cluster partition model. After obtaining the volatility sequence V by the GARCH model, Fisher's optimal dissection method is then applied to cluster volatility.

4.4. Results of volatility cluster partitions

The cluster statistic can be used to identify partition points as described in Section 3.1. To illustrate the cluster algorithm, as an example, we select the latest one year data with 252 daily observations in 2018. Panel A of Figure 1 shows the cluster statistic for every partition point. Cluster statistics peak at $(197, 0.0012)$ which suggests that the optimal partition point is the 197-th day in the sample, or October 12, 2018. This result is consistent with the pattern in Panel B of Figure 1, where volatility jumps to a high value after October 12. In contrast, volatilities before October 12 mostly stay in relatively low values, though two peaks, located at 23-th and 79-th points in Panel A of Figure 1, exist and they represent two sub-optimal partition points. The corresponding dates for these two peaks in Panel B of Figure 1 are February 5 and April 26, respectively. Although these two points can also split the sample into two distinguishing parts, they are not as effective as the optimal partition point corresponding to October 12.

5. Applications for VaR

Risk control is indispensable in investment and financial management. Value-at-risk (VaR) provides a risk metric for managing the risk of the asset portfolios of financial institutions. Returns of financial assets or portfolios are usually assumed to follow a certain distribution, and the mean and variance of that distribution are estimated over a fixed window. Many VaR estimation methods involve the GARCH model. To simplify VaR calculation and to improve precision, we apply constant volatilities with cluster partitions as described in Section **??**. We calculate the VaR and compare it with the conventional methods based on the GARCH models and Regime-switching models.

5.1. Estimations by the GARCH model

GARCH is an important model for stationary time series, and stationary assumptions are required. Before using the GARCH model to predict the volatility of the return

series, we first conduct the stationarity test for the return series and the ARCH effect test on return residuals. ~~We then use the Kupiec test, a classical unconditional covering test suggested by Kupiec (1995), to test VaR validity.~~

Parameters of GARCH(1,1) model are estimated and reported in Table 2. Figure 2 shows the volatility curve predicted by the GARCH(1,1) model. The figure exhibits clear volatility clustering and hence, it is appropriate to use the cluster method to split the volatilities in the full sample.

5.2. Determine cluster number

In applying the volatility cluster partition model, we need to determine the optimal cluster number and partition points. Figure 4-6 show graphic method, slope statistic and information statistic. The loss function in Figure 4 suggests that the loss gets smaller as the cluster number increases. The value of the loss function decreases slowly after $N = 50$. As N increases beyond this point, the value of the loss function approaches zero. For brevity, we omit $N > 800$ in the figure. However, the graphic method can only show the optimal cluster number roughly. The slope statistic $S(N)$ denoted in (**??**) is shown in Figure 5. As $S(N)$ approaches 1, the cluster number is closed to optimal. However, from this graph it is difficult to identify a concrete point to be an optimal cluster number. Thus, the optimal cluster number is still fuzzy.

These two statistic methods can only express the optimal cluster number figuratively. Information based statistic in (17)is plotted in Figure 6. Although the two statistic methods above can express the optimal cluster number figuratively, it is hard to provide an accurate number of partitions. In addition, as the cluster number increases, too many pieces of clusters may lead to an over-fitting problem. To overcome these difficulties, we use the information-based statistic in (17) to come up with a measure of fitness on the model. From Figure 6, the statistic in (17) reaches the minimum at $N = 345$. For a bigger cluster number $N > 345$, the increasing value of the statistic indicates more information loss due to the complexity of model selection.

5.3. Results of volatility cluster partitions

For each cluster number, we use GARCH cluster partirion model to calculate the optimal split points and fit volatility. Figures 7-9 present the estimated volatility with cluster partitions at $N = 500$, $N = 345$ and $N = 100$, respectively. For a large cluster

number, e.g., $N = 500$ in Figure 7, the volatility obtained by GARCH cluster partirion model appears to be quite closed to the dynamic volatility infered by the GARCH model in Figure 2. As cluster number decreases, estimated volatilities are more smoothed. For small partition numbers, e.g. $N = 100$ in Figure 9, the graph seems much simpler. In despite of that top volatilities still resemble white noises, the cluster with lower volatilities becomes more expressive of volatility clustering phenomenon.

6. Conclusion

In this paper, we propose a volatility cluster partition model based on Fisher's optimal dissection method to forecast return volatility. Our objective is to find a model that not only can explain the volatility behavior well, but also provides a better tool for volatility forecasts. We find that our cluster partition model generates much higher accuracy in forecasting return volatility than any conventional models and offers a much more reliable tool for risk management using the VaR approach. Moreover, applying the volatility cluster partition method to the Black-Scholes implied volatility significantly improves the forecast for option prices. When decomposing volatility into short- and long-run components, applying the volatility cluster partition method to the long-run component improves the estimation and forecasting performance of the GARCH-MIDAS model. The volatility cluster partition method provides more reliable volatility forecasts to better estimate option prices, which can facilitate efficient arbitrage for traders in the derivatives market and improves market efficiency. In addition, as volatility is an important factor for predicting returns (see Ang et al. (2006) and Chung et al. (2019)), better volatility forecasts by the cluster partition method facilitate asset pricing tests and risk management.

Appendix A. Forward algorithm for optimal clusters

In this appendix, we explain how we calculate the optimal cluster by a forward algorithm. For a given ordered data $V = (v_1, v_2, \cdots, v_n)$ and cluster number N, we have a classifier π with corresponding loss function $L(\pi) = \sum_{k=1}^{N} D(C_k)$, dividing V into clusters $(C_1, \ldots, C_N)$, within each cluster C_k the subscript of the first data is I_k. Denote the optimal classifier as $B(n, N) = \arg\min_{\pi} L[\pi(n, N)]$. By dynamic programming

principal (DPP) method, The optimal classifier satisfies

$$\begin{aligned} L(B(n,N)) &= \min L[P(n,N)] \\ &= \min_{I_N}\{L(B(I_N-1,N-1)) + D(C_N)\}, \end{aligned}$$

where I_N is the subscript of the N-th split point and the first data in the last cluster C_N. This suggests that once the last group C_N is determined, the other series $(v_1, \cdots, v_{I_N-1})$ classified into $N-1$ groups should also be optimal. This process satisfies until the classifier becomes a binary classification.

$$\begin{aligned} L(B(I_N-1,N-1)) &= \min_{I_{N-1}}\{L(B(I_{N-1}-1,N-2)) + D(C_{N-1})\}, \\ &\vdots \\ L(B(I_3-1,2)) &= \min_{I_2}\{D(C_1) + D(C_2)\}. \end{aligned}$$

Generally, a cluster includes at least several elements. Let h be the given minimum number involved in a cluster. By the recursive relations above, the algorithm searching the optimal partition points can be summarized as follows.

Step 1. Starting from $m = h+1$ to $m = n$, for each m we search the optimal partition point $I_2(m)$ of the following function by the golden section method.

$$L(B(m,2)) = \min_{I_2(m)}\{D(C_1) + D(C_2)\},$$

where $C_1(m) = \{x_{I_1}, \ldots, x_{I_2(m)-1}\}$, $C_2(m) = \{x_{I_2(m)}, \ldots, x_m\}$ and $I_1 = 1$. Further, we search the optimal position m_{opt} of m, such that the corresponding set of $\{L(B(m,2))\}_{m=h+1}^{n}$ reaches its minimum. Concurrently, we record I_3, I_2, C_1 and C_2 as $I_3 = m_{opt}$, $I_2 = I_2(m_{opt})$, $C_1 = C_1(m_{opt})$ and $C_2 = C_2(m_{opt})$, respectively.

Step 2. Suppose that we have obtained $k-1$ $(4 < k < N)$ optimal cluster partition points, $I_1, I_2, \ldots, I_{k-1}$. From the recursive relation

$$L[B(m,k)] = \min_{I_k}\{L(B(I_k-1,k-1)) + D(C_k)\},$$

the next step we need to take is to search the optimal $D(C_k)$. By the similar method to Step 1, we can obtain the optimal C_k and I_k.

Step 3. Let $k = k + 1$ for $k < N$. Repeat Step 2 until $k = N$. Then N optimal clusters are achieved.

References

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

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., 2009: Uncertainty across volatility regimes. Journal of Applied Econometrics, 34(3):437-455.

Asgharian, H., Hou, A.J., Javed, F., 2013: The importance of macroeconomic variables in forecasting stock return variance: a GARCH-MIDAS approach. Journal of Forecasting, 32(7):600-612.

Bai, J., Perron, P., 1998: Estimating and testing linear models with multiple structural changes. Econometrica, 47-78.

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

Baker, S.R., Bloom, N., Davis, S.J., Kost, K., 2019:Policy news and stock market volatility. NBER Working Paper.

Barndorff-Nielson, O. E. and Shephard, N., 2002: Estimating quadratic variation using realised variance. Journal of Applied Econometrics.

Black, F., Scholes, M., 1973: The pricing of options and corporate liabilities. Journal of Political Economy, 81(3):637-654.

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.

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

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.

Carrasco, M., Hu, L. Ploberger, W., 2004: Optimal Test for Markov Switching. Society for Economic Dynamics. 374.

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.

Chow, G.C., 1960: Tests of equality between sets of coefficients in two linear regressions. Econometrica: Journal of the Econometric Society, 591-605.

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

Chuong, L., Nikolai, D., 2018: Forecasting of realised volatility with the random forests algorithm. Journal of Risk & Financial Management, 11(4):61.

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

Conrad, C., Loch, K., 2015: Anticipating long-term stock market volatility. Journal of Applied Econometrics, 30(7):1090-1114.

Corsi, F., 2008: A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics, 7(2):174-196.

Cox, J.C., 1975: Notes on option pricing I: Constant elasticity of variance diffusions. Unpublished note, Stanford University, Graduate School of Business.

Cox, J.C., Ross, S.A., 1976: The valuation of options for alternative stochastic processes. Journal of Financial Economics, 3(1-2):145-166.

Duan, J.C., 1995: The GARCH option pricing model. Mathematical Finance, 5(1):13-32.

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., Ghysels, E., Sohn, B., 2013: Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics, 95(3):776-797.

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.

Fama, E.F., 1965: The behavior of stock-market price. The Journal of Business, 38(1):34-105.

Fisher, W.D., 1958: On grouping for maximum homogeneity. Journal of the American Statistical Association, 53(284):789-798.

Garcia, R., 1998: Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models. International Economic Review, 39: 763-788.

Giacomini, R., White, H., 2006: Tests of Conditional Predictive Ability. Econometrica, 74, 1545-1578.

Glosten, L. R., Jagannathan R., and Runkle, D. E., 1993: On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48:1779-801.

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.

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

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

Heston, S.L., 1993: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies, 6(2):327-343.

Hull, J., White, A., 1987: The pricing of options on assets with stochastic volatilities. Journal of Finance, 42(2):281-300.

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

Jorion, F., 1996: Risk^2 : Measuring the risk in value at risk. Financial Analysts Journal, 52(6):47-56.

Klaassen, F.J.G.M., 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.

Kyong, J.O., Kim, T.Y., Min, S., 2005: Using genetic algorithm to support portfolio optimization for index fund management. Expert Systems with Applications, 28(2):371-379.

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.

MacQueen, J. , 1967: Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.

Mandelbrot, B., 1967: The variation of some other speculative prices. Journal of Business, 40(4):393-413.

Meddahi, N., 2002: A theoretical comparison between integrated and realized volatility. Journal of Applied Econometrics, 17(5): 479-508.

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.

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

Poterba, J.M., Summers, L.H., 1984: The persistence of volatility and stock market fluctuations. The American Economic Review, 75(5):1142-1151.

Pérignon, C., Smith, D.R., 2008: A new approach to comparing VaR estimation methods. Journal of Derivatives, 16(2):54-66.

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

Rubinstein, M., 1994: Implied binomial trees. The Journal of Finance, 49(3):771-818.

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.

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.

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

Table 1: Results of ARCH effect test

Hypothesis H_0: there is no ARCH effect on residual terms				
α	Test rejection	p-values	Test statistic	Critical value
1%	True	0.000	352.93	6.6349
5%	True	0.000	352.93	3.8415
10%	True	0.000	352.93	2.7055

Note: The parameter α is the significant level. "Test rejection = True" indicates the rejection of hypothesis H_0 holds. From both p-values and test statistics are greater than critical values at different levels, it indicates that the evidence is strong to reject the null hypothesis.

Table 2: Estimation result of GARCH(1,1) model

Conditional probability distribution: Gaussian			
Parameter	**Value**	**Standard error**	**Statistic**
Constant	0.00	0.00	5.07
GARCH{1}	0.90	0.00	200.57
ARCH{1}	0.08	0.00	22.74

Note: In this table, the model GARCH(1,1) is estimated and the statistic of its parameters are list out. GARCH{1} denotes the parameter estimation of one-lagged conditional variance. ARCH{1} denotes the parameter estimation of one-lagged sample variance.

Table 3: Results of Kupiec test on VaR under GARCH Cluster Partition model

Test for 2 years			**Test for 10 years**		
Confidence level	Non-rejection region	Fails	Confidence level	Non-rejection region	Fails
99%	$1 < n < 10$	12	99%	$16 < n < 36$	56
95%	$16 < n < 36$	24	95%	$104 < n < 148$	140
90%	$37 < n < 64$	41	90%	$222 < n < 282$	236

Note: In this table, under GARCH Cluster Partition model we show the results of Kupiec test on the failures of VaR at different significant level. Both the non-rejection region and the failure numbers are presented under corresponding significant level. Here, the 2-year data sample of returns of S&P 500 index starts from January 3, 2017 to December 31, 2018. The 10-year sample covers the return data of S&P 500 index from January 3, 2009 to December 31, 2018. The number n is the acceptance failure number in Kupiec test. If the failure number falls into the range of non-rejection region, it implies that the Kupiec test is validated and the employed method can be acceptable to calculate VaR. For instance, in the test for 2 years, it shows that the fail number, 12, is above the range of $1 < n < 10$ at confidence level, 99%, which suggests that the method we use underestimates VaR values. Similarly in the sample of 10 years at 99% confidence level, the fail number, 56, fall above the upper bound of 36, which means it underestimates the falis and the risk. Besides, failure numbers at other confidence levels fall into the ranges of non-rejection regions for both 2-year and 10-year sample.

Table 4: Results of Kupiec test on VaR under variance-covariance method

Test for 2 years			**Test for 10 years**		
Confidence level	Non-rejection region	Fails	Confidence level	Non-rejection region	Fails
99%	$1 < n < 10$	18	99%	$16 < n < 36$	59
95%	$16 < n < 36$	37	95%	$104 < n < 148$	133
90%	$37 < n < 64$	43	90%	$222 < n < 282$	204

Note: In this table, under variance-covariance method we show the results of Kupiec test on the failures of VaR at different significant level. Both the non-rejection region and the failure numbers are presented under corresponding significant level. Here, the 2-year data sample of returns of S&P 500 index starts from January 3, 2017 to December 31, 2018. The 10-year sample covers the return data of S&P 500 index from January 3, 2009 to December 31, 2018. In addition, the number n is the acceptance failure number in Kupiec test. If the failure number falls into the range of non-rejection region, it implies that the Kupiec test is validated.

Table 5: Results of Kupiec test on VaR under GARCH(1,1) model

Test for 2 years			**Test for 10 years**		
Confidence level	Non-rejection region	Fails	Confidence level	Non-rejection region	Fails
99%	$1 < n < 10$	15	99%	$16 < n < 36$	61
95%	$16 < n < 36$	30	95%	$104 < n < 148$	148
90%	$37 < n < 64$	47	90%	$222 < n < 282$	247

Note: In this table, under GARCH(1,1) model we show the results of Kupiec test on the failures of VaR at different significant level. Both the non-rejection region and the failure numbers are presented under corresponding significant level. Here, the 2-year data sample of returns of S&P 500 index starts from January 3, 2017 to December 31, 2018. The 10-year sample covers the return data of S&P 500 index from January 3, 2009 to December 31, 2018. In addition, the number n is the non-rejection failure number in Kupiec test. If failure number falls into the range of non-rejection region, it implies that the employed method can not be rejected to calculate VaR.

Table 6: Results of Kupiec test on VaR under EGARCH(1,1) model

Test for 2 years			**Test for 10 years**		
Confidence level	Non-rejection region	Fails	Confidence level	Non-rejection region	Fails
99%	$1 < n < 10$	15	99%	$16 < n < 36$	46
95%	$16 < n < 36$	31	95%	$104 < n < 148$	119
90%	$37 < n < 64$	45	90%	$222 < n < 282$	200

Note: In this table, under EGARCH(1,1) model we show the results of Kupiec test on the failures of VaR at different significant level. Both the non-rejection region and the failure numbers are presented under corresponding significant level. Here, the 2-year data sample of returns of S&P 500 index starts from January 3, 2017 to December 31, 2018. The 10-year sample covers the return data of S&P 500 index from January 3, 2009 to December 31, 2018. In addition, the number n is the non-rejection failure number in Kupiec test. If failure number falls into the range of non-rejection region, it implies that the employed method can not be rejected to calculate VaR.

Table 7: Results of Kupiec test on VaR under RSGARCH model

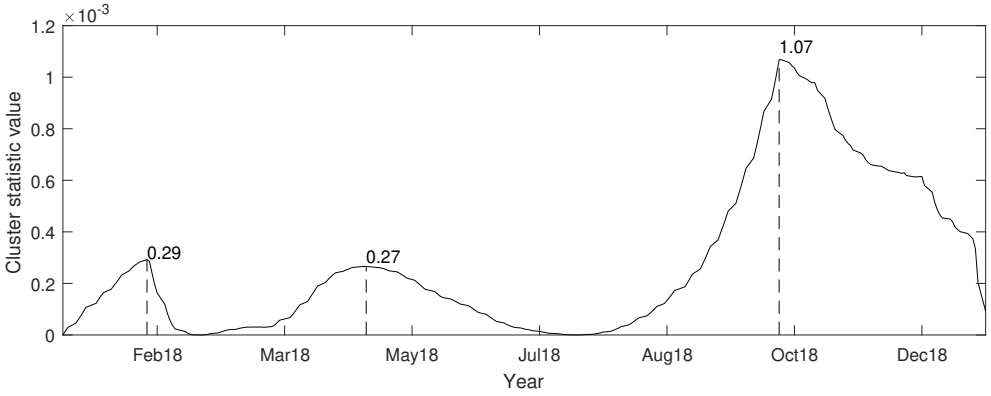
Test for 2 years			**Test for 10 years**		
Confidence level	Non-rejection region	Fails	Confidence level	Non-rejection region	Fails
99%	$1 < n < 10$	28	99%	$16 < n < 36$	111
95%	$16 < n < 36$	41	95%	$104 < n < 148$	210
90%	$37 < n < 64$	57	90%	$222 < n < 282$	292

Note: In this table, under RSGARCH model we show the results of Kupiec test on the failures of VaR at different significant level. Both the non-rejection region and the failure numbers are presented under corresponding significant level. Here, the 2-year data sample of returns of S&P 500 index starts from January 3, 2017 to December 31, 2018. The 10-year sample covers the return data of S&P 500 index from January 3, 2009 to December 31, 2018. In addition, the number n is the non-rejection failure number in Kupiec test. If failure number falls into the range of non-rejection region, it implies that the employed method can not be rejected to calculate VaR.

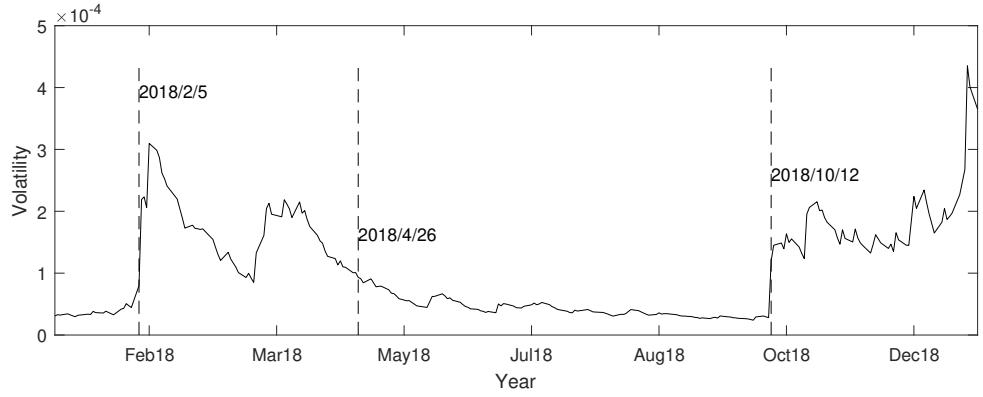
Table 8: Comparison of Kupiec test and accuracy on VaR

Panel A: Test for 2 years							
Method	99.5%	99%	97.5%	95%	92.5%	90%	Accuracy
Historical simulation	0.09	1.51	**4.87****	**3.68***	**5.33****	**5.72****	94.65%
Variance-covariance	**36.48*****	**20.35*****	**12.9*****	4.42**	0.53	1.2	94.42%
GARCH	**18.75*****	**10.92****	**2.94***	0.15	0.96	1.57	95.56%
EGARCH	**28.97*****	**20.35*****	**9.88*****	2.39	0.79	0.07	94.42%
RS-GARCH	**18.75*****	**10.92*****	**3.85****	1.36	0.16	1.20	94.16%
GARCH Cluster partition	**12.78*****	**8.91*****	**2.94***	0.61	0.05	0.62	95.33%
Panel B: Test for 10 years							
Method	99.5%	99%	97.5%	95%	92.5%	90%	Accuracy
Historical simulation	0.44	0.13	0.79	0.12	0.00	0.09	96.14%
Variance-covariance	**50.29*****	**33.35*****	**10.59*****	0.32	0.94	**10.63*****	95.80%
GARCH	**58.22*****	**38.70*****	**20.06*****	**2.95***	1.67	0.19	95.21%
EGARCH	**63.71*****	**52.32*****	**40.84*****	**11.76*****	1.67	0.09	94.92%
RS-GARCH	**40.36*****	**29.96*****	2.25	0.09	1.45	**10.63*****	95.99%
GARCH Cluster partition	**38.00*****	**26.71*****	**9.86*****	**3.91****	0.39	1.24	95.47%

Note: In this table, we mainly summarize and compare the statistics of Kupiec test and the accuracy of different methods. The 2-year data sample covers from January 3, 2017 to December 31, 2018. The 10-year sample covers January 3, 2009 to December 31, 2018. Column 2-8 reports the statistics of Kupiec test that eximines whether the failure rate is equivalent to the pre-specified VaR level. The bold numbers indicate the rejection of the null hypothesis at the level of 1% with ***, 5% with ** and 10% with *, respectively. The "Accuracy" is a ratio with respect to events without failure over the total events. The best model is the one with the least rejections and the highest accuracy. These results show that the VaRs estimated by cluster partition model are credible, better than other models.



Panel A. Cluster statistic



Panel B. Volatility

Figure 1: Cluster statistic on corresponding volatility. In this figure, we select the returns of S&P500 index in the year, 2018, as our empirical data. In Panel A, we show the cluster statistic discussed in section 3.1 for every partition point. Cluster statistics peak at $(197, 0.0012)$, which suggests that the optimal partition point is the 197-th day in the sample, or October 12, 2018. This result is consistent with the phenomenon in Panel B, where volatility jumps to a high value after October 12. Panel B reports corresponding volatilities. Three most significant cluster statistics in Panel A correspond to three partition points for volatility in Panel B.

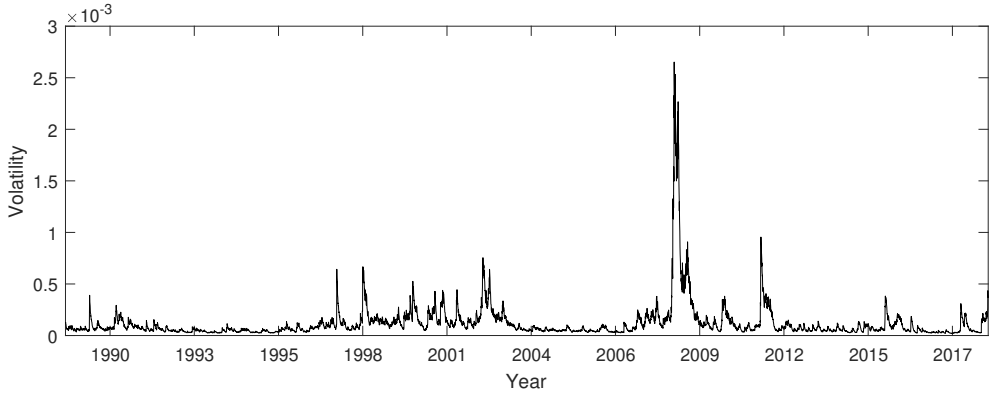


Figure 2: Volatility estimated by GARCH(1,1) model. To see volatility clusters, in this figure we select the returns series of S&P 500 index with long time period from January 3, 1989 to December 31, 2018. In this graph, we can see several obvious volatility clusters. For instances, from the year 1992 to 1996 the volatility keeps stable at a low level. Then, from 1997 to 2003 the volatility turns to a higher level than before and still keeps stable during this period.

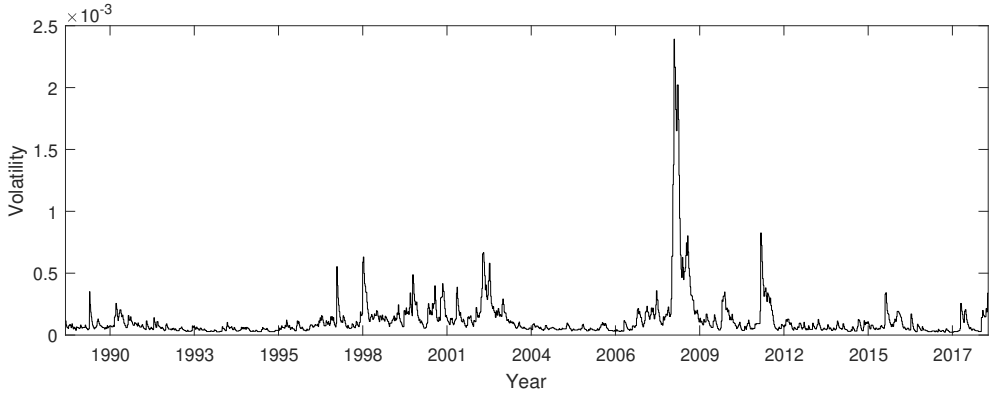


Figure 3: Cluster volatilities with 1000 clusters. From this graph, we can see that cluster volatilities with 1000 clusters is similar to volatilities estimated by GARCH(1,1). It implies that as the partition number N approaches to the sample size n, the volatility obtained by the cluster partition model would be similar to that obtained by the GARCH(1,1) model. This implies that the volatility estimated by GARCH(1,1) is the limit of the volatility estimated by the cluster partition model at the partition number $N = n$.

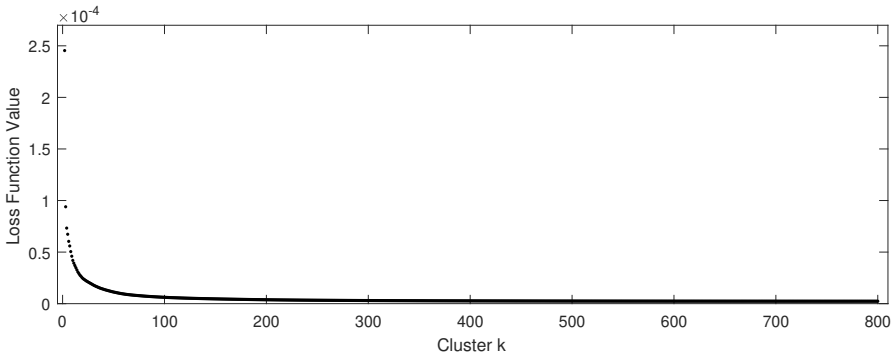


Figure 4: Loss function value. In this graph, we use the same data sample from January 3rd, 1989 to December 31th, 2018. The horizontal axis denotes the cluster numbers employed. From this graph, we can see that the loss function defined in (**??**) decreases monotonically and approaches zeros as cluster number increases.

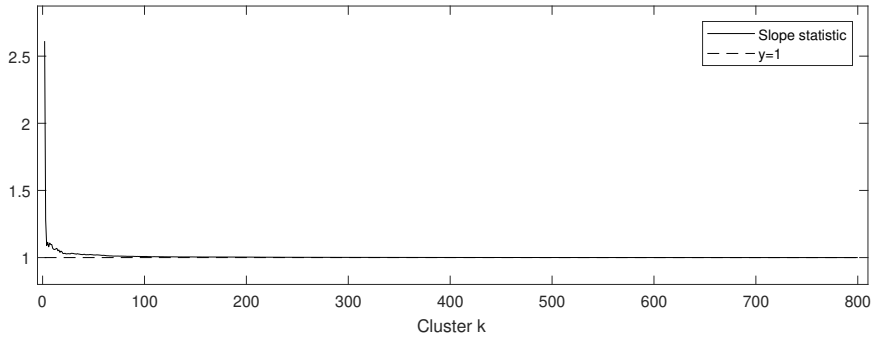


Figure 5: Slope statistic of $S(N)$. The slope statistic in (**??**) is shown and the data sample is the same as that in Figure 4. As $S(N)$ approaches 1, the cluster number is closed to optimal. However, in this graph the slope statistic never traverses y = 1, so that it is very hard to give a concrete point to be an optimal cluster number. Thus, the optimal cluster number is still fuzzy.

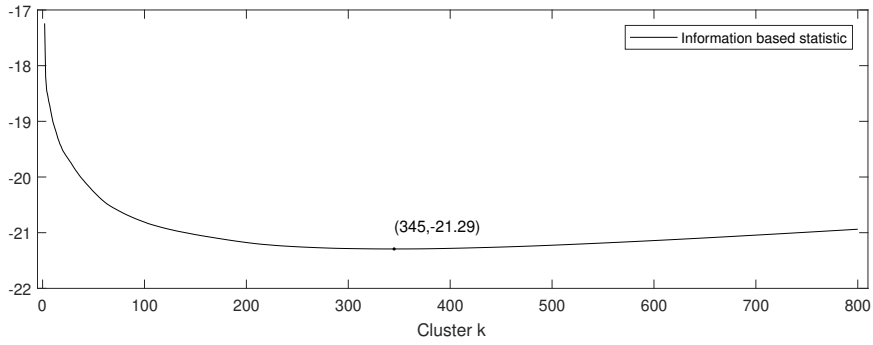


Figure 6: Information based statistic $\psi(N)$. In this graph, the information based statistic defined in (17). It decreases as cluster number increases at first, but increases when cluster number reaches $N = 345$. So, by this statistic, we can see that the optimal cluster number is $N = 345$.

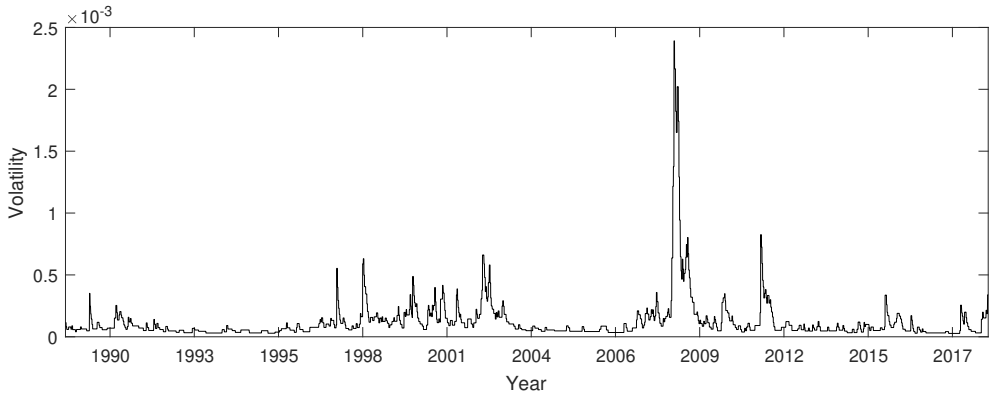


Figure 7: Cluster volatilities with $N = 500$. The horizontal axis denotes the year from 1989 to 2018. In each cluster, volatility is a constant and the curve is plat. The whole shape of this figure is similar to that of volatilities estimated by GARCH(1,1) model.

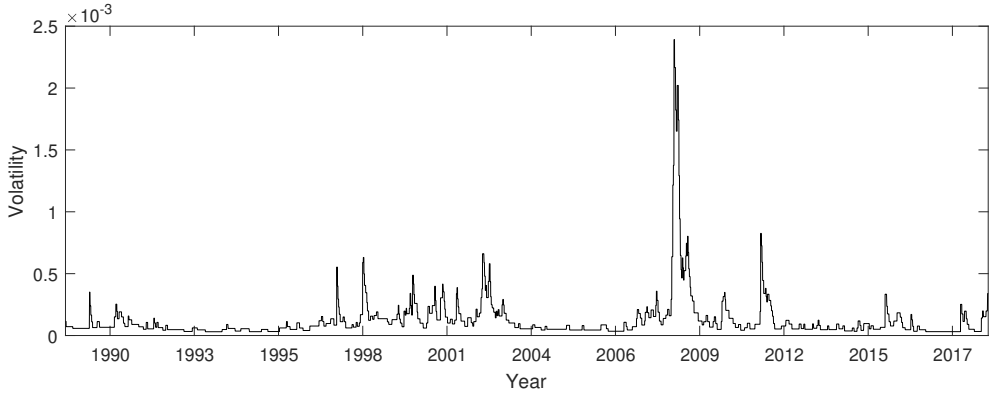


Figure 8: Cluster volatilities with $N = 345$. Compared this graph with the Figure 7, lower volatilities of this graph are more representative. But from the view of visual point, there is no obvious difference between Figure 7 and Figure 8. When the value of N gets larger, the figure gets closer to GARCH(1,1), and the degree of difference becomes less obvious. So $N = 345$ could be a good representative as a large partition number.

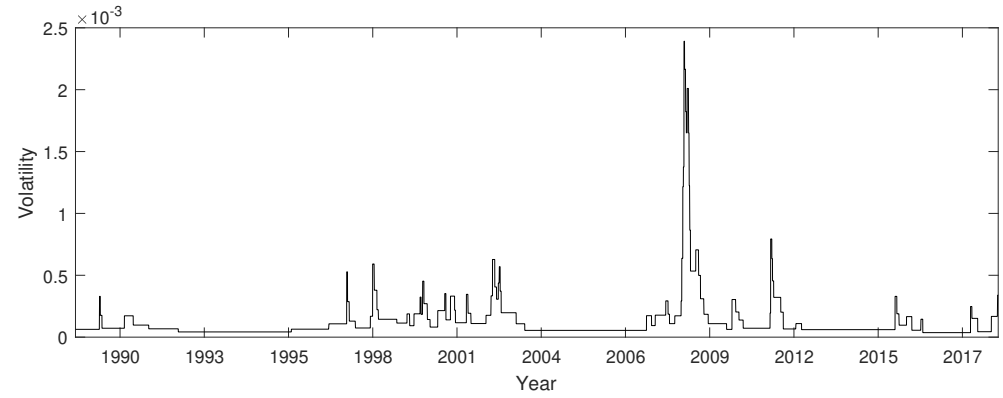


Figure 9: Cluster volatilities with $N = 100$. This graph seems much simpler than Figure 7 and Figure 8. In despite of top volatilities still like white noises, the cluster with lower volatilities becomes more expressive of volatility clustering phenomenon.

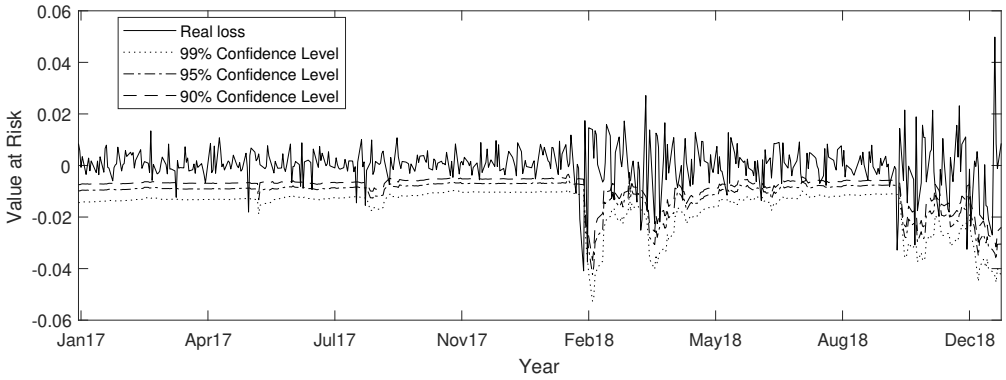


Figure 10: Value at risk estimated by GARCH Cluster Partition model. In this graph, VaRs under 99%, 95% and 90% confidence levels are estimated by volatility cluster partition method, respectively. The estimation window is from January 01, 2017 to December 31, 2018. For each day of the estimation window, we apply GARCH Cluster Partition model to predict the volatility of next day. When real loss penetrates the line of VaR, failure number counts 1.

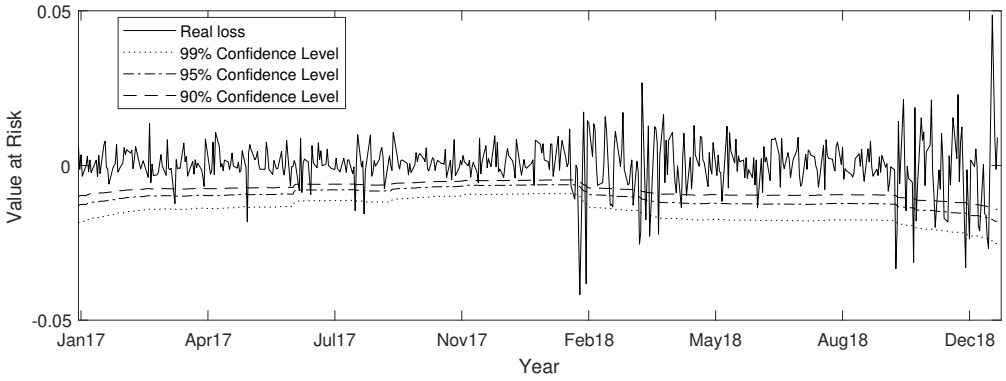


Figure 11: Value at risk estimated by variance-covariance method. In this graph, VaRs under 99%, 95% and 90% confidence levels are estimated by variance-covariance method, respectively. When real loss penetrates the line of VaR, failure number counts 1. VaR estimated by variance-covariance method is relatively stable and thus cannot respond to the market fluctuation immediately.

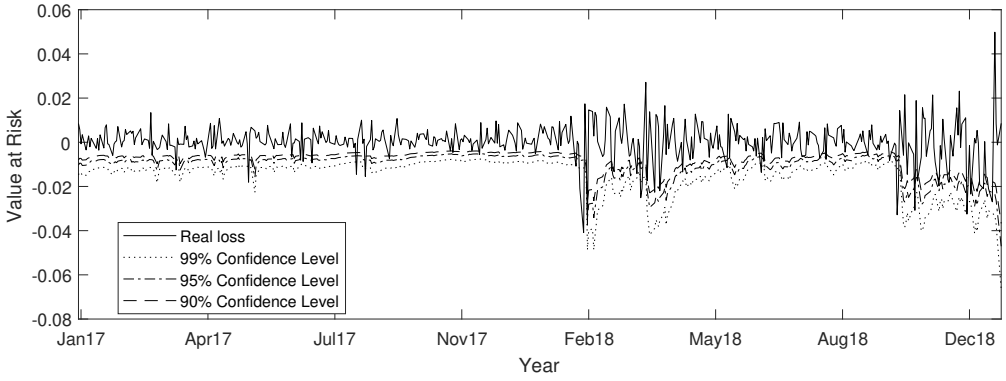


Figure 12: Value at risk estimated by GARCH(1,1) model. In this graph, VaRs under 99%, 95% and 90% confidence levels are estimated by GARCH(1,1) model, respectively. When real loss penetrates the line of VaR, failure number counts 1. VaR estimated by GARCH(1,1) model is sensitive to market fluctuations and has a quick response whereas VaR is sometimes overestimated.