Statistical Research on Macroeconomic Big Data: Using a Bayesian Stochastic Volatility Model

Minglei Shan(✉)

Shandong Youth University of Political Science, Jinan 250103, China

Abstract. The alternative variation of variance in Stochastic Volatility (SV) models provides a big data modelling solution that is more suitable for the fluctuation process in macroeconomics for de-scribing unobservable fluctuation features. The estimation method based on Monte Carlo simula-tion shows unique advantages in dealing with high-dimensional integration problems. The statis-tical research on macroeconomic big data based on Bayesian stochastic volatility model builds on the Markov Chain Monte Carlo estimation. The critical values of the statistics can be defined exactly, which is one of the drawbacks of traditional statistics. Most importantly, the model pro-vides an effective analysis tool for the expected variable generation behaviour caused by macroe-conomic big data statistics.

Keyword: Bayesian stochastic volatility models · Economic big data statistics · Monte Carlo simulation algorithms

Owing to the impact of the COVID-19 epidemic, China's macroeconomic performance is currently exhibit-ing more multi-dimensionality and heterogeneity. In this case, the lag and limitations in the application of tradi-tional statistical theories and research tools are more no-ticeable. Therefore, modeling application methods based on time-varying fluctuation processes are valued and men-tioned by academic scholars and government economics officials working in the area of macro-statistics. In view of the new characteristics of volatility clustering and nonlin-ear dynamic structure presented by the time series in the economic system during the epidemic period, the modeling analysis of the Bayesian stochastic volatility model grad-ually showed its advantages in fusion and algorithm.

1 Theoretical Basis and Research Significance

1.1 Related Concepts

The Bayesian algorithm, developed by the English mathematician Thomas Bayes in the 18th century, assumes the prior knowledge used to estimate the parameters are random variables, where unknown parameters are independent of the distribution. Based on

Z. Deze et al. (Eds.): BDTA 2020/WiCON 2020, LNICST 371, pp. 90–103, 2021.
https://doi.org/10.1007/978-3-030-72802-1_7

the probability rule, when the probability of the sample event is close to the overall probability, the posterior information of the parameter is obtained. The basic formula is:

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)} \propto p(X|\theta)p(\theta),$$

In this formula, P(θ|X) represents the posterior probability density function of the parameter given for the sample X, P is a constant function of a relationship formed between the posterior probability of θ when X occurs, and x represents the prior probability density function of the parameter.

The stochastic volatility model, also known as the SV model in statistics, was proposed by the mathematicians Taylor and Shephard in the 1980s and 1990s. It was originally used to explain the autoregressive phenomenon of serial volatility in economics (performance conditional heteroscedasticity, also called the GARCH model). Random fluctuations focus on measuring the degree of random variation in a particular time series. Random volatility is defined as the standard deviation or covariance in a continuous difference model. Its model has the characteristics of dynamic fluctuations, assuming that the interference item is unobservable and follows the process of random fluctuations.

The Monte Carlo Method is used to construct a sample with a π(x) stationary distribution through probability theory and statistical theory, and perform various statistical inferences based on these π(x) samples. The core of the algorithm is to use the experimental mathematical method of digital simulation in order to construct or describe the probabilistic process by mastering the geometric quantity and geometric characteristics of the object motion. Based on the probability model, the simulated experimental sample is used as an approximate solution to find and establish an estimate quantity, and then perform statistical inference.

The nonlinear structure of macroeconomics. Nonlinear structure refers to the characteristics that a node may have multiple successors and several precursors. Random sampling and empirical research show that the linear expression of macroeconomics is not often the case and is mostly non-linear. The interaction of macroeconomic factors, the process of information internalization to form market prices, and the operation of economic fluctuations are all inherently non-linear. Research in these areas has inaugurated a new field for macroeconomics research and put forward higher requirements for most macroeconomic modeling.

1.2 Research Significance

Theoretical significance: The final goal of the development of SV Model modeling theory is to apply aggregate volatility observations, by also including unobservable implicit volatility variables. In the entire development process of the time-varying volatility model, various typical characteristics of the macroeconomic time series can be effectively portrayed and the similarity function can be accurately expressed, thus conforming to the driving force of actual modeling to the greatest extent. Owing to the limitation of the level of macroeconomic models, the volatility observation of the time-varying Monte Carlo estimation method is slower than that of the GARCH model. The SV Model modeling theory can describe more accurately the characteristics of volatility, and has more important significance in predicting the big data of uncertainty continuous coefficient change point fluctuation and volatility risk.

Practical significance: The COVID-19 epidemic has severely affected the economic and social operations, to such a degree that it caused large economic downturns and turmoil within a certain period of time and made the non-linear structure of the macro economy obvious. Moreover, the instability is expected to grow further, which overall shows a decline at the macroeconomic level. The trend has strengthened, economic growth momentum has weakened, and overall volatility has increased. For similar non-linear states, the traditional Kalman filter method cannot perform a comprehensive analysis. It is necessary to consider the approximate filter method in the form of Bayesian estimation model for processing. The full use of the characteristics of prior information is beneficial to the actual situation and big data analysis and forecasting.

2 Purpose and Content

2.1 Main Research Purpose

This research aims to focus on a typical time-varying volatility model: SV Model, which introduces a new stochastic process with strong fitting ability and high estimation difficulty. Since the model contains unobservable latent wave variables, it is difficult to obtain an accurate expression of the likelihood function for its nonlinear structure. In addition, the extended form of nonlinear structure in macroeconomic big data is more complicated, making it difficult to estimate potential state variables and related parameters. Therefore, model estimation is traditionally presented as the focus and difficulty in the modeling process. The time-varying volatility model based on steady-state simulation is a frontier subject of modern econometrics research and has important theoretical value. Sequential Bayesian filtering technology has strong adaptability to macroeconomic fluctuation state filtering and parameter learning methods, and can better deal with the estimation of large-scale nonlinear non-Gaussian state space models. This model can effectively separate expected information, accurately carry out big data statistical prediction and analysis, and promote the development of big data algorithms in the field of volatility modeling. This kind of research has good adaptability to the expansion of various volatility models. Furthermore, in-depth discussion of its algorithm regarding the improvement and application of the nonlinear structure state of macroeconomic big data, as well as big data related modeling based on volatility model research field has high reference value.

In the entire development trajectory of the big data time-varying volatility model, we can effectively describe the various typical characteristics of the macroeconomic time series. This is the driving force for the continuous development and change of the model form. The basic goal of the Monte Carlo estimation method of the time-varying volatility model in the field of macroeconomic big data statistics is to describe more accurately the characteristics of various fluctuations, which are the risk management of the macroeconomic field. In addition, it has an important role in the prediction of macroeconomic risks and price fluctuations. Traditional statistical inference methods are based on the overall data, and pay more attention to the derivation of causality. This impedes them in meeting the requirements of big data operations in terms of calculation amount and sensitivity. The big data simulation technology recursively updates the estimated value, avoiding the storage and reprocessing of the previous observation data, and effectively reducing the actual calculation cost. Moreover, the introduction of big data simulation technology into the estimated value of the time-varying volatility model can effectively solve the problems of repeated calculation and time lag in parameter estimation and state prediction, and realize the time-varying sequential prediction effect of the online data intelligent volatility model.

With the development of modern econometrics and computer technology, the non-linear structural time-varying fluctuation process modeling method is provided as a powerful analysis tool for effective macroeconomic risk management, and its use has been partially applied to asset portfolios and macro capital flows field. It can be seen from the existing literature that the construction and estimation of time-varying volatility models have been rapidly developing. Especially in recent years, the application of statistical forecasting has gradually become a hot spot in the field of model design and academic research. In addition, the time series data of macroeconomic big data usually does not meet the independent repeated test conditions in classic statistics. Therefore, with the continuous development of macroeconomic big data statistical research, the observation and calculation techniques for the behavior of expected variables should not only change accordingly, but improve. In the Bayesian method, the model parameters are random variables, which have a specific statistical distribution form. This proves to be an effective tool for solving the macroeconomic nonlinear structure observation problem. Research in this area mainly focuses on the classical statistical modeling theory.

The distribution of parameter estimates and test statistics in the volatility models are unknown, so it is difficult to determine the level of accurate critical values. Owing to the complexity of the model form, it is difficult to obtain analytical expressions of model parameters. In an attempt to describe the volatility characteristics of the time series of big data more accurately and comprehensively, people have developed various extended forms of time-varying volatility models. The Monte Carlo model provides a set of effective estimation methods for various complex models, especially for those that are difficult to estimate. Owing to the high correlation between economic big data samples, traditional methods have slow convergence speed, which is not conducive to empirical analysis. However, the design of effective estimation methods for various extended time-varying volatility models and related application research still needs in-depth work. Since this model is a typical nonlinear and non-Gaussian state space model, it only has the analytical expression of the posterior distribution in concept, so it is necessary to find the approximate solution of Bayesian estimation. When new observations appear in this process, the posterior probability density of the Monte Carlo estimation method will be re-estimated based on the Markov chain mantissa algorithm. Further use of the estimation technology based on the Bayesian stochastic volatility model is used to convert the state of the complex system into the prior probability density of the simulated and predicted state. Following, the latest observations are used to make corrections to obtain the posterior probability density, and, finally, to obtain the best estimate of the state. In other words, it is not necessary to process all the data of the past time every time, but to estimate the current state vector based on the observation vector of the current time and the state vector of the previous one or more times. This technology can also be applied to online data analysis that is common in macroeconomic analysis.

2.2 Research Context and Technical Route

Based on the analysis of the research background and literature review, the research group used induction and simulation methods to sort out the theory of the Bayesian stochastic volatility model. A macroeconomic big data statistical program based on this model was further proposed, and an online reasoning method was explored. According to the current research status, the following main issues are discussed:

Under the classical statistical modeling theory system, research of improved algorithms and applications of regression are conditional on heteroscedasticity models and their respective extended models in the area of macroeconomics. Furthermore, they are dependent on the accurate critical values of parameter estimation and test statistical distribution in the volatility model. With the development of the macroeconomic system, time series data in this field no longer meet the independent repeated test conditions in classical statistics, and the behavior of expected variables will also change accordingly. The Bayesian model has a specific statistical distribution form, where the parameters present random variables. The framework of the Bayesian method is used to analyze the time-varying fluctuation model, which provides a broad exploration space for solving the above problems.

For the application of the Bayesian model in the non-linear structure of the macroeconomics, we focus on big data algorithms. First, in order to develop an extended form of the time-varying volatility model, we use the auxiliary particle filter algorithm corresponding to the Bayesian sequence to accurately and comprehensively describe the volatility characteristics of the macroeconomic time series. Owing to the complexity of the model form, it is difficult to obtain analytical formulas for model parameters. The Monte Carlo simulation method provides an effective calculation method for various complex models, especially the SV Model, which is difficult to estimate. Owing to the high correlation between data samples, the traditional Markov chain Monte Carlo (MCMC) method has a very slow test speed, which is not conducive to empirical analysis. Therefore, it proves to be a challenge to design effective MCMC estimation methods for various extended forms of time-varying volatility models and conduct related application research.

The MCMC sampling algorithm's estimation of the SV model will generate potential state variables of the simulated data. Therefore, the corresponding MCMC sampling method will generate multiple sampling chains. It can be seen from the data calculation of the sequence autocorrelation function of each sampling chain that there are two single steps. The state autocorrelation performance of the MCMC sampling method declines very slowly, while the state of the combined sampling method where the Bayesian model participates in the sample size autocorrelation performance declines quickly, and maintains a low level of correlation. This is mainly due to the joint sampling process, where the latent state variables are extracted as a whole, so that the Markov chain has a faster convergence speed. Besides the fact that the joint sampling method based on normal approximation has a large deviation from the state value, other methods have little difference in the estimation accuracy of state variables. In each algorithm, calculating the mean square error (MSE) between the estimated value of the state variable and the true value leads to obtaining a more accurate value.

2.3 The Main Content of the Research

Owing to the instability of the macroeconomic system, venture capital and macroeconomic control tools with volatility as their main component have received extensive research and attention. With the development of "descriptive economics", big data modeling and analysis of market price volatility under macroeconomic regulation have gradually become the focus of attention in both theoretical and practical circles. In view of the new characteristics of time series and changes in economic aggregates (quantity space) in the macro economy, the big data modeling method of time-varying fluctuations has been greatly developed. The SV Model in the time-varying volatility model has a strong fitting ability and a more challenging estimation difficulty.

SV Model contains unobservable latent wave variables, which makes it difficult to obtain an accurate expression of the likelihood function. Its various expansion forms are more complicated, and it is difficult to estimate potential state variables and parameters. Therefore, the estimation of big data models has always been the key issue in the modeling process. With the continuous development of computing technology, the estimation method based on the Monte Carlo simulation has shown its unique advantages in dealing with high-dimensional integration problems. Among them, the MCMC algorithm has become the fastest growing and most widely used model method among time-varying volatility model estimation methods. Combined with the application of background in macroeconomics, the research group made an effort to: improve the SV Model, study the Bayesian reasoning process, design the MCMC sampling algorithm based on the Bayesian model, and compare the effectiveness of various MCMC sampling algorithms in the SV Model. In the application field of big data models, two time-varying volatility models are used to study the dynamic relationship between the inflation level and the uncertainty persistence coefficient change point during the fluctuation period of the macroeconomic continuous coefficient change point, and under the influence of global epidemic conditions. The market trend of China can provide a useful reference for the application of this model in the current macroeconomic risk management and economic policy formulation. In this way, we follow the standard Bayesian conjugate prior to setting the model (Alston, Mengersen, and Pettitt, 2012), and select the hyper-parameters to correspond to the non-informative prior settings. This way the information contained in the data set corresponds to the appropriate covariance conditions, where weighting coefficients are assigned to the MCMC prior distribution:

$$p(\boldsymbol{\rho}) = \mathrm{Dir}\left(\boldsymbol{\rho}; \alpha_1^{(0)}, \ldots, \alpha_K^{(0)}\right);$$

The prior distribution of the mean conditioned on the covariance matrix is an independent multivariate normal distribution:

$$p(\mu \,|T) = \prod_{j=1}^{K} N_d\left(\mu_j; m_j^{(0)}, \left(\beta_j^{(0)} T_j\right)^{-1}\right);$$

The prior of the exact matrix is given by the MCMC distribution:

$$p(T) = \prod_{j=1}^{K} W\left(T_j; v_j^{(0)}, \Sigma_j^{(0)}\right);$$

Therefore, the joint distribution will eventually be: p(y, z, θ) = p(y, z | θ)p(ρ)p(μ | T), where the number of p(T) are all hyper-parameters.

In the MCMC estimation method, each time a new observation is obtained, the posterior probability density must be re-estimated, with all previous samples needing to be retained, which may take up a lot of the calculation. The Sequential Monte Carlo (SMC) technology uses the system transition model to predict the prior probability density of the state, and then uses the latest observation data to obtain the posterior probability density of the state, thereby calculating the optimal estimate of the state. Meaning, the technology can be used to analyze online data commonly found in macroeconomics and economic analysis.

The focus of the research is an SV MODEL and its extended SMC estimation method. A sequential Bayes filter parameter learning algorithm is proposed, building upon the existing parameter learning method based on artificial noise process. Following, various filtering algorithms are compared and analyzed. Finally, due to the impact of emergencies in the macroeconomic field, the fluctuations of big data time series are usually clustered at different levels. Establishing a model with variable structural fluctuation characteristics has important practical value for improving the accuracy of macroeconomic market fluctuation forecasts and avoiding macroeconomic investment risks. Therefore, the research team focused on the variable structure form of the SV MODEL. In particular, the spotlight was on the comparative analysis of the structure expansion form of the big data model and the sequential Monte Carlo algorithm, combined with the current domestic macroeconomic actual continuous coefficient change point shock situation. The model was used for the purposes of potential fluctuation prediction and its application in emergency detection showed good performance.

The research is based on the estimation of the time-varying volatility Bayesian model of Monte Carlo simulation method, including the focus on several topics:

The research group proposed a variety of standard MCMC sampling algorithms for SV MODEL based on an analysis of the research status and related background of time-varying wave models at home and abroad, as well as Monte Carlo simulation methods. The SV MODEL introduces random error terms into the wave equation to make the data modeling process more flexible and the estimation process more difficult. The method of the simulation analysis and comparison is used to systematically summarize the combined the design of the standard SV MODEL and MCMC algorithm, which further promotes the research progress of the SV MODEL and highlights its basic statistical characteristics. Based on several main estimation methods of the SV MODEL in recent years and by analyzing their respective advantages and disadvantages, the sampling algorithm of a standard SV MODEL is clarified.

Based on the background of macroeconomic applications, the existing SV MODEL has been expanded and improved. Namely, the improved MCMC sampling algorithm and several important extensions of the long memory SV MODEL are studied, as well as the Bayesian reasoning process. Furthermore, the design of the corresponding MCMC sampling and macroeconomic application background improves the model. In the category of multivariable SV MODEL, the Gibbs joint sampling algorithm is used to design the model by long memory in order to infer a special form of forecast distribution, and further determine the difference between the asymmetric mean of the SV MODEL, and the current macroeconomic level and uncertain fluctuations of the dynamic relationship. A Bayesian comprehensive national credit premium model based on multi-factor SV MODEL is established to distinguish the characteristics of average recovery rates of different scales. Moreover, a multi-step MCMC method based on mixed normal distribution is used to simulate credit premium index sequences for companies using different remaining payment conditions and perform applicability analysis. In view of the reasoning requirements of online data and the fact that SV MODEL and its extended form are nonlinear non-Gaussian state-space models, together with the time-varying and uncertainty of the current economic level, the known parameters of the sequential Monte Carlo model are systematically discussed using the SMC technology. The Monte Carlo model in SV MODEL performs big data simulation analysis. The specific calculation method is as follows:

The distribution $q(\theta, z)$ is chosen to minimize the Kullback-Leibler (KL) divergence between the approximate density $q(\theta, z)$ and the true joint density $p(\theta, z|y)$. For this reason, an attempt is made to obtain a relatively low lower limit of the marginal density $p(y)$. In order for the maximization method to be used to estimate the parameters of the objective approximation function, the correlation density must be manipulated and re-expressed to allow for the introduction of a variational approximation function (McGrory and Titterington, 2007). From then on, the calculation model can be derived for a simulation analysis of parameters:

$$\begin{aligned}\log p(y) &= \log \int \sum_{\{z\}} p(y, z, \theta) d\theta \\ &= \log \int \sum_{\{z\}} q(\theta, z) \frac{p(y, z, \theta)}{q(\theta, z)} d\theta \\ &\geq \int \sum_{\{z\}} \log \frac{p(y, z, \theta)}{q(\theta, z)} d\theta\end{aligned}$$

Among them, the multi-step MCMC method of mixed normal distribution can analyze the compound state model through the Bayesian method. When $q(\theta, z) = p(\theta, z|y)$, it is completely minimized, $q(\theta, z)$ is close to the true density, and $Q(\theta, z)$ is limited to the factorized form $Q(\theta, z) = q\ \theta(\theta)\check{Z}(z)$.

In online data analysis, where the model parameters are unknown, the parameter learning method based on sequential Monte Carlo is further discussed. As a result, a sequential Bayes filter parameter learning method is processed by using artificial noise technology. After adopting a parameter learning method based on the SMC model and having sufficient statistical characteristics, the statistical model variables are fully considered. The filter density function is decomposed to employ the sequential Bayes filter parameter learning algorithm. On the one hand, the parameter learning method based on auxiliary particle filter and its improved algorithm are studied in the simulation analysis. On the other hand, the effectiveness of this algorithm is illustrated by comparing various filtering algorithms. Therefore, the simulation results of the sequential Bayes filter parameter learning algorithm can be further discussed and compared with the existing learning algorithm. It is found that the standard SV is a more practical and effective alternative method.

For the variable structure of an SV MODEL, the model contains some extreme cases that are of particular importance to risk management. The research group relies on the sequential Bayes filter parameters for simulation and empirical analysis of the learning algorithm. The results show that this algorithm can effectively describe the dynamic structural characteristics of macro-market fluctuations and can avoid the subjective bias of prior information. The use of the sequential Bayes filter parameter learning algorithm is threefold: 1) to conduct empirical research on China's macroeconomics, 2) to decompose the joint filter density function from another angle, 3) and to conduct empirical research on the variable structure characteristics of macroeconomic fluctuations. It can

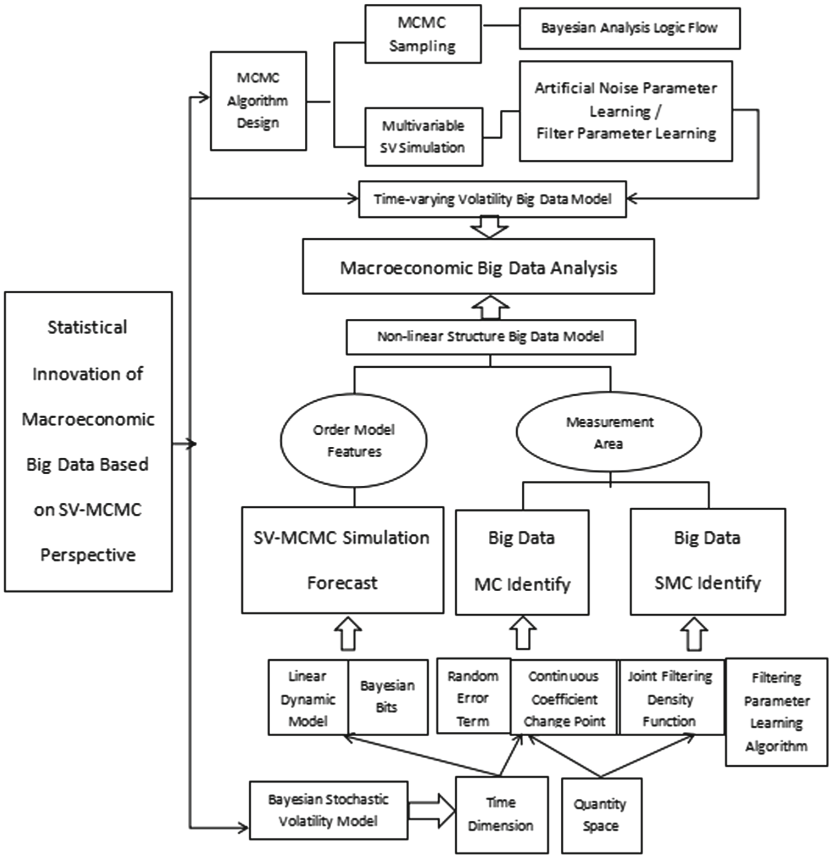

Fig. 1. Research on macroeconomic big data statistics using a Bayesian stochastic volatility model

be said that the use of the SV MODEL in the MCMC algorithm can lead to improvements in MC identification and calculation in terms of model prediction, and emergency detection estimation efficiency and accuracy under potential fluctuations.

The overall framework and specific steps of the research content are shown in Fig. 1:

3 Analysis

The Bayesian stochastic volatility model is used to identify and analyze the SMC of major macroeconomic big data indicators. At present, major foreign research institutions such as the National Bureau of Economic Analysis and the Oxford Institute for Economic Research are using the Bayesian method to make statistics on the macroeconomic situation. Compared with traditional volatility modeling and estimation methods, the Bayesian method has the advantage of maintaining effective estimation capabilities in situations where the affected continuous coefficient change point fluctuation is greater than the number of continuous fluctuations at the model stopping point. The research group draws on the Arch model measurement method proposed by the American scholar Engel, a Nobel Prize winner in economics, and relies on the metrological stochastic volatility model for processing and research. The unit of continuous volatility is set to (Yt), and the sum is calculated according to the MCMC sampling algorithm. The posterior distribution is:

$$Y_t = \left[\int_0^1 Y_{jt}^{\frac{1}{\lambda_{f,t}}} dj \right]^{\lambda_{f,t}}, \quad 1 \leq \lambda_{f,t} \leq \infty$$

The time is labelled as jt, the number of continuous wave points is presented as dj, and the function curvature caused by the wave is λf, t.

Since a single variable structure prediction method is suitable for different systems with limited measurement accuracy, the model should have different parameter selection mechanisms in order to be able to automatically select the best prediction parameters in the same system. In practice, the period of macroeconomic fluctuations is getting shorter and shorter, and the information response is becoming more and more sensitive, resulting in a large amount of online data. For time-varying market conditions, the MCMC sampling simulation technology has a great dependence on the model and parameter settings. Every operation in this estimation process is based on the overall data, which is abundant. From the perspective of variable diagnosis structure points, segmented modeling is a relatively simple method. Also, the maximum likelihood value of the estimation is obtained based on sequential Monte Carlo simulation. In these equations, Yt is replaced by ŷi, where ŷi corresponds to the weighted observation and is defined as follows:

$$\hat{\boldsymbol{y}}_i = \frac{\gamma_i \times \boldsymbol{y}_i}{\frac{1}{n}\sum_{i=1}^{n} \gamma_i}$$

The expected value required to update the expression remains unchanged. In specific, it keeps the same form adopted by the standard VB algorithm. This algorithm outlines the pseudo-code of the weighted VB algorithm, which we call Core Set Variation Bayes (CVB). The CVB algorithm input is expressed as:

$$C = \{(\gamma(\times 1), \times 1), (\gamma(\times 2), \times 2), \ldots, (\gamma(\times N), \times N)\}$$

In such way, it allows for a hierarchical structure of data points to be constructed and the weight of the multiplicity of the sampling points to be associated with the log likelihood. If the estimated log likelihood has the smallest variance, the weight is set to the optimum. This kind of structure of the sampling set helps to provide higher probability for observations that are far away from the initial cluster center. Furthermore, the sampling deviation can be fixed by adjusting the weights related to the sampling probability. This finally allows for a possibility to build an appropriate core set from the entire data set based on the weight.

In the process of time series modeling, the main purpose of the research is to extract the statistical characteristics rules from the observation data containing noise, with a final goal to make effective statistical inferences for MC identification. Compared with traditional time series models, the state-space big data modeling process can express the dynamic structure of observed variables and latent state variables, thus directly reflecting the movement of the data generation process. More importantly, in the framework of state space modeling, time series models (such as ARMA model and SV MODEL) can be integrated into a unified structure in such a way that estimation methods with general characteristics can be used for analysis. When compared with the MCMC, the standard SV method has higher calculation accuracy, although there may be cases where the analysis needs to be performed in a shorter time frame. For super large data, this would mean facing the problem of cumbersome calculations. A resolution to this can be proposed by reducing the amount of data that actually has to be processed in the first place. Deleting a part of the data set before analysis is a lot less incommoding when the data sets provide the same or very similar information. For example, consider trying to fit a mixed model to an image. We do not have to analyze all of the observed intensity levels present in the complete data set to get a good mean estimate for that particular cluster.

In addition to resolving the cumbersome calculation issue, the state space form of this big data model also provides a more effective way to simplify the maximum likelihood estimation and deal with missing values. Since the system identification of the state space model is based on the Bayesian principle, it mainly reflects to the approximate filtering of the state space model. In the field of frequency statistics, the sequential Monte Carlo is chosen for maximum likelihood estimation method. For the estimation of the linear Gaussian state space model, statistical inference and analysis can be completed through the iterative process of mean and variance. In a statistical sense, the Kalman filter can be used as the best estimation method. However, for the nonlinear non-Gaussian state-space models, the approximation results are very sensitive to the resampling process. This is mainly because the Bayesian estimation of the model is more complicated or has a high dimension, and therefore cannot be applied to online data estimation, nor can it be analyzed in a broad sense.

Consider using other optimization methods to solve the maximum likelihood estimation problem, such as approximate filtering methods or simulation-based filtering methods. Knowing the algorithm parameters, the regional metric distribution can be obtained through the following iterative process. The distribution formula is as follows:

$$P\left(x,\ |y'\theta\right) = p(x, Ix, \theta)p(y, Ix, \theta)$$

Once the distribution function P(x,|y'θ) in the model with known parameters is obtained, that is, θ is a constant factor, the problem is transformed into a filtering analysis of the model state variables. The joint posterior density function of the state variables I = N, with sampling the increase of the number N follows the core idea of a Monte Carlo estimation.

This formula cuts in from the mixed smooth distribution, combined with the filtering process, and proposes an algorithm for sequential Bayes filter parameter learning. In order to obtain the sampling form of direct filtering, we consider the filtering process and statistical marginal distribution of state variables. The basic idea of the approximate filtering method is to transform the nonlinear non-Gaussian state space model into a linear Gaussian model. The Kalman filter method is used to realize the system identification, and the recognition field is represented by the extended Kalman filter (EKF) and the Kalman filter. However, this method only plays a role in approximate sampling. In addition, Gaussian mixture distribution approximation is used to deal with the estimation problem of the state space model. In practical applications, not all models can be converted into a Gaussian linear form. The approximate filtering method with models is used to estimate state variables and parameters. In particular, these methods may have common problems. The integral approximation process is based on the approximation result, which in return drives the approximation error to be larger and larger, making the estimated value to deviate from the actual value. This method is shown in Fig. 2:

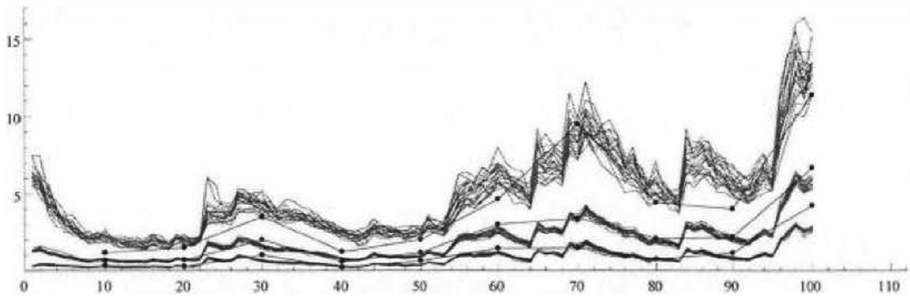

Fig. 2. Approximate filtering method used to estimate fluctuation characteristics of the SV model order model

In nonlinear and non-Gaussian state space big data models, simulation-based filtering methods provide a more general solution for the system identification problem. This method is based on Monte Carlo simulation, combined with sequential Bayesian filtering. It should be noted that the sampling distribution of the parameters is set in the form of a mixed normal distribution. When new data is received, changes in the model are reflected in time. Nevertheless, the effective information can be extracted to correct and estimate the model parameters. Compared with the MCMC method, this method has multiple advantages: high sensitivity, small calculation amount, high estimation accuracy, and higher practical significance. A combination of increasing number of observed variables, rapid growth of computing power and the continuous decrease of computing costs provides a broad development space for the convergence and complexity of the SMC method in the state space of the known parameter random fluctuation model.

4 Conclusion

In the field of macroeconomic risk research, the modelling and estimation of big data time series are undoubtedly of great significance. With the development of computer technology and Bayesian method, big data modelling overcomes the problem of difficulty to accurately determine the critical value of test statistics in the traditional statistical modelling process. In addition, with the continuous development of the economy and the macroeconomic system, Monte Carlo simulation can help solve the complex numerical calculation problems in the estimation of big data models, and the behaviour of expected variable generation will also change accordingly. The time-varying fluctuation model based on Monte Carlo simulation also provides an effective research tool for solving the above stated problems. Owing to the complexity of the time-varying wave model and its model in extended form, it is difficult to obtain analytical expressions of big data model parameters using only MCMC simulation method and SMC simulation method. The Monte Carlo simulation method provides a set of effective estimates for various models. The application of SV Model is expected to have more research and contributions in the future, especially in the field of macroeconomics that is characterised by many wave points and complex forms.

Fund Project. Major project of statistical research project of Shandong Province in 2019: "statistical research on macro-economic big data based on Bayesian stochastic fluctuation model: from the perspective of Monte Carlo simulation" (No.: kt1907); supported by the growth plan of young teachers in Shandong Province.

References

1. Tanaka, M., Katayama, T.: Robust Kalman filter for linear discrete-time system with Gaussian sum noises. Int. J. Syst. Sci. **18**(9), 1721–1731 (2019)
2. Chow, H.K.: Robust estimation in time series: an approximation to the Gaussian sum filter. Commun. Stat. Theory Methods **23**(12), 3491–3505 (2014)
3. Kahn, H., Marshall, A.W.: Methods of reducing sample size in Monte Carlo computations. J. Oper. Res. Soc. Am. **1**(5), 263–271 (2013)
4. Kloek, T., van Dijk, H.K.: Bayesian estimations of equation system parameters: an application of Monte Carlo integration. Econometrica **46**(1), 1–20 (2018)
5. Shephard, N.M., Pitt, K.: Likelihood analysis of non-Gaussian measurement time series. Biometrika **84**(3), 653–667 (2007)
6. Richard, J.F., Zhang, W.: Efficient high-dimensional importance sampling. J. Econ. **141**(2), 1385–1411 (2017)
7. Liya, H.: Bayesian stochastic volatility model based on Monte Carlo simulation and its application research. Hunan University (2011)

8. Wang, J., Xiao, F., Chen, J., Yao, X.: Reliability analysis of pile foundation based on Bayesian theory and Monte Carlo simulation. J. Undergr. Space Eng. **13**(S1), 85–90 (2017)
9. Meng, L., Zhang, S.: Empirical eigenfunction method for parameter estimation of SV model. Syst. Eng. **22**(12), 92–95 (2004)
10. Su, W., Zhang, S.: Multiple long memory SV model and its application in Shanghai and Shenzhen stock markets. J. Manag. Sci. **7**(1), 38–44 (2004)
11. Xu, M., Zhang, S.: Research on estimation method of long memory stochastic volatility model based on wavelet transform. China Manag. Sci. **14**(1), 7–14 (2006)