ARTICLE IN PRESS

Economic Modelling xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/econmod

Calculating Value-at-Risk for high-dimensional time series using a nonlinear random mapping model

Heng-Guo Zhang[a,b], Chi-Wei Su[a], Yan Song[c], Shuqi Qiu[c], Ran Xiao[d], Fei Su[d,*]

[a] *Department of Finance, Ocean University of China, Qingdao, Shandong, China*
[b] *Department of Finance and Economics, Shandong University of Science and Technology, Jinan, Shandong, China*
[c] *College of Information Science and Engineering, Ocean University of China, Qingdao, Shandong, China*
[d] *Finance Discipline Group, UTS Business School, University of Technology, Sydney, Australia*

ARTICLE INFO

JEL:
C32
C45
C53

Keywords:
Extreme learning machine
High-dimensional space
Value-at-Risk
Random mapping
GARCH model
Time series

ABSTRACT

In this study, we propose a non-linear random mapping model called GELM. The proposed model is based on a combination of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and the Extreme Learning Machine (ELM), and can be used to calculate Value-at-Risk (VaR). Alternatively, the GELM model is a non-parametric GARCH-type model. Compared with conventional models, such as the GARCH models, ELM, and Support Vector Machine (SVM), the computational results confirm that the GELM model performs better in volatility forecasting and VaR calculation in terms of efficiency and accuracy. Thus, the GELM model can be an essential tool for risk management and stress testing.

1. Introduction

There has been intensive research on risk modeling to develop more sophisticated risk management techniques and models. Following the increasing uncertainty in financial markets and the global financial crises since the 1990s, it is of great interest to practitioners, regulators, and academic researchers (Angelidis et al., 2004). Value-at-Risk (VaR) has emerged as a suitable and remarkable tool to quantify risk, and became popular and prevalent during the 1990s because of its simplicity and easy implementation (Giot and Laurent, 2004). VaR proposed by Jorion (1995) is defined as the worst expected loss over a given horizon at different levels of confidence. Even though conceptually simple and usable, VaR estimations have become more complicated and sophisticated during the last decade.

*Overall, the VaR estimation methods can be categorized into three major types (*Kim and Lee, 2016*).* The first one is the parametric method, wherein a return process is generated with error terms that follow a specific distribution. A popular example is the Generalized Autoregressive Conditionally Heteroskedastic (GARCH) model. The GARCH model was first proposed by Bollerslev (1986), which uses historical variance to predict future volatility under the assumption of conditional heteroskedasticity. A number of extensions to the GARCH model have been proposed and applied to financial markets, including the integrated GARCH (IGARCH) model introduced by Engle and Bollerslev (1986), the exponential GARCH (EGARCH) model proposed by Nelson (1991), etc. In this study, we extend the literature on conditional heteroskedasticity to encompass a broader class of GARCH processes, namely, a non-parametric GARCH-type model. A detailed introduction to GARCH-type models is given in Section II.

The second type is the semi-parametric method. It applies either the quantile regression approach proposed by Koenker and Bassett (1978) or extreme value distribution-based methods, for example, Extreme Value Theory (EVT) introduced by Smith (1989). The EVT is well known for modeling extreme tails by analyzing the upper and lower quantiles of the corresponding distribution.[1] For example, Wang et al. (2012) use the composite quantile regression (CQR) method as proposed in Zou and Yuan (2008), to estimate the intermediate quantiles, and then extrapolated these estimates to calculate the extreme quantiles using the EVT.

The third type includes the non-parametric method, which is especially useful for analyzing big data, as in the case of machine learning. An artificial neural network (ANN) is one of the most

* Corresponding author.
E-mail address: fei.su@uts.edu.au (F. Su).

[1] For a thorough introduction to EVT, see Smith (1989) and McNeil (1997), as well as the references therein.

http://dx.doi.org/10.1016/j.econmod.2017.02.014
Received 21 July 2016; Received in revised form 7 January 2017; Accepted 15 February 2017
0264-9993/

Please cite this article as: Zhang, H.-G., Economic Modelling (2017), http://dx.doi.org/10.1016/j.econmod.2017.02.014

primitive machine learning frameworks, motivated by the learning capabilities of the human brain, and capable of performing parallel computation for time series forecasting. There are many ANN variations, such as back propagation (BP) networks introduced by Williams and Hinton (1986), and the radial basis function (RBF) proposed by Lowe (1989), among many others. However, bottlenecks in areas such as overfitting, local minima, and computation time, etc, can restrict the scalability of these models in conventional implementations. Cortes and Vapnik (1995) put forward the statistical learning theory and develop the Support Vector Machine (SVM) based on the structural risk minimization (SRM) principle, which seeks to minimize an upper bound on the Vapnik Chervonenkis (VC) dimension of the generalization error. Beyond these models, deep learning has become a popular topic since Hinton and Salakhutdinov (2006) proposed the deep belief networks (DBN). A DBN can handle complex, high-dimensional time series by exploiting multilayer hierarchical neural network architectures. In addition, convolutional neural networks (CNNs) introduced by LeCun et al. (1989) are a family of multilayer neural networks that automatically learn hierarchical features, and therefore extract complex translation and distortion invariant features in higher layers. A brief introduction to these models of relevance is presented in Section II as well. One of the most important advantages of these machine learning models is that they are data driven. In other words, the models can be reasonably estimated with no prior assumptions on the nature of the error distribution.

In the era of big data, how to model and calculate the VaR for high-dimensional time series becomes a challenging topic. Berman (2013) finds that the unique features of big data are high dimensionality, complexity, massiveness, heterogeneity, incompleteness, noisiness, and erroneousness. Thus, conventional approaches to measuring risks face major challenges in handling big data. For example, most conventional methods used to calculate the VaR were originally designed for relatively small data sets, and may require additional assumptions on the distribution of the error terms (e.g., a normal distribution assumption). In addition, Chen and Zhang (2014) argue that when analyzing big data, it is essential to speed up the estimation procedure, especially for conventional approaches, in order to reduce the computation time and memory requirements.

Recently, the extreme learning machine (ELM) has attracted increasing attention. It provides a greater generalization performance with a faster learning speed. The ELM proposed by Huang et al. (2006) was originally developed for single-hidden layer feedforward networks (SLFN), rather than using the classic gradient-based algorithms, which has been extended to the generalized SLFN. In general, Huang et al. (2012) deem that the essence of ELM is that when the input weights and hidden layer biases are randomly assigned, the output weights can be computed using the generalized inverse of the hidden layer output matrix. In this study, motivated by the remarkable success of ELM in terms of generalization performance and learning speed, we propose a non-linear random mapping model called GELM, which is based on a combination of the conditional heteroskedasticity processes and ELMs. More specifically, we apply an ELM algorithm to high-dimensional time series of GARCH processes, and thereby extend the multivariate GARCH process in a non-parametric framework.

The main characteristics of the GELM model are as follows. Firstly, compared with parametric and semi-parametric models, the GELM model utilizes a random mapping method, which does not employ the Gaussian likelihood to estimate the parameters. Specifically, the GELM randomly generates the hidden node parameters, for which it keeps the same virtues of the random parameters as in the ELM model. Once the input weights and the hidden layer biases are randomly assigned, the output weights of the GELM model can be calculated analytically using the simple generalized inverse operation of the hidden layer output matrix.

The second characteristic lies in the fact that the proposed GELM model is a non-linear data-driven model. More specifically, the GELM model learns and approximates the dynamics of time series in a non-linear fashion directly from the data, with no prior assumptions. That is, in the GELM framework, it is not to assume that the conditional dependencies are all contained in the conditional mean and variance, and the variables are not necessarily independent over time. For example, the model functions well, even when the stationary and ergodic conditions are not satisfied. Thus, the GELM model is well suited for complex problems with little prior knowledge, but with a large number of observations.

Thirdly, the GELM model is more noise tolerant with regard to connections between the system state variables. In other words, in the learning process with the given data, the GELM model can always correctly infer the hidden part of the time series, even if the sample contains noisy information. It can be estimated under weak or even no prior assumptions on the error distribution. In addition, the model can learn and estimate complex systems with incomplete, non-linear, and non-stationary data structures. Thus, it can handle big data sets in high-dimensional spaces.

The rest of the paper is organized as follows. In Section 2, previous works on GARCH-type models and feedforward neural networks are reviewed. The basic ELM model is introduced in Section 3. Section 4 proposes the GELM model and Section 5 reports on the experimental results, and discusses related issues. Lastly, Section 6 concludes the paper, with possible directions for future research.

2. Literature review

Traditional time series tools for the conditional means such as autoregressive moving average (ARMA) model introduced by Box and Jenkins (1976) have been extended to analogous models for higher moments of time series. Autoregressive Conditional Heteroscedasticity (ARCH) models are commonly used to estimate and forecast changes in the second moment of financial time series.[2] Since the seminal work of Engle (2002), much progress has been made in understanding GARCH models and their multivariate extensions.

1) *The univariate GARCH*. Given an univariate GARCH model on return series, we can infer the conditional distribution of the return, and thereby calculate the VaR of a long or short position. The univariate GARCH type models can be classified into four types.

   The first type is the basic GARCH model proposed by Bollerslev (1986), which can be written as:

$$\sigma_t^2 = c + \sum_{i=1}^{n} \alpha_i \xi_{t-i}^2 + \sum_{i=1}^{n} \gamma_i \sigma_{t-i}^2 \tag{1}$$

   where σ_t^2 denotes the conditional variance, α_i and γ_i are non-negative ARCH and GARCH coefficients respectively, and ξ_t is the random error. The integrated GARCH model proposed by Engle and Bollerslev (1986) is a special case of the GARCH model, where the sum of the ARCH and GARCH coefficients equals one:

$$\begin{cases} \sigma_t^2 = \sum_{i=1}^{n} \alpha_i \xi_{t-i}^2 + \sum_{i=1}^{n} \gamma_i \sigma_{t-i}^2 \\ \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \gamma_i = 1 \end{cases} \tag{2}$$

   The second type relaxes the assumption of symmetry in the conditional variance specification. For example, the most commonly adopted asymmetric GARCH model is the GJR model developed by Glosten et al. (1993):

[2] For a survey of ARCH-type models, please see Bollerslev et al. (1992), Bera and Higgins (1993), Pagan (1996), among many others.