

# GARCH Parameter Estimation by Machine Learning

Tetsuya Takaishi

**Abstract**— It is of great importance to estimate volatility of asset returns for risk management in empirical finance. The GARCH model is often used to estimate volatility. To utilize the GARCH model, we need to estimate model parameters so that the model matches the underlying return time series. Usually the maximum likelihood or the Bayesian method is used for the parameter estimation of the GARCH model. In this study we apply the machine learning technique for the parameter estimation. We minimize the loss function defined by the likelihood function of the GARCH model. The minimization is done by the Adam optimizer of TensorFlow. We find that the machine learning estimates the model parameters correctly. We also investigate the convergence property of the Adam optimizer and show that the convergence rate increases as the learning rate increases up to a certain maximum learning rate. Over the maximum value, the minimization fails with the optimizer.

**Index Terms**— GARCH model, Likelihood Function, Machine Learning, Bayesian Inference

## I. INTRODUCTION

In empirical finance, volatility or variance of asset returns is important for risk management. Since volatility is latent in return time series, one needs to use a certain volatility estimation technique. A popular technique is to use models that mimic properties of return time series. The most commonly used models for volatility estimations in empirical finance are the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model [1,2,3]. The GARCH model can successfully capture some known properties of returns time series such as volatility clustering, fat-tailed return distribution, insignificant autocorrelation in returns and so forth. These properties are now classified as “Stylized facts” [4].

There exist many extended GARCH models that are designed to refine the model by capturing further empirical properties of returns. For instance, the exponential [5], GJR [6], Asymmetric [7,8] and Rational [9] GARCH models are constructed to capture the asymmetric property of volatility. It is well known that while stock returns show the asymmetry in volatility, for exchange rates and Bitcoin [10] returns the asymmetry in volatility is less significant.

To utilize these GARCH-type models we need to estimate model parameters so that the model matches the return time series. Usually the maximum likelihood method or the Bayesian estimation performed by the Markov Chain Monte Carlo method [11-18] is used for the GARCH parameter estimations. In this study we use the alternative path, namely the machine learning for the parameter estimations. In the machine learning process, we minimize a loss function

defined by the likelihood function of the GARCH model. Then we show that the parameter estimation by the machine learning can also lead to the correct values of parameters.

## II. GARCH MODEL

The GARCH model [1,2,3] we use in this study is defined as follows. Let  $r_t$   $\{t=1, \dots, N\}$  be a return at time  $t$ .  $r_t$  is given by

$$r_t = \sigma_t \varepsilon_t, \quad (1)$$

where  $\sigma_t^2$  is the volatility and  $\varepsilon_t$  the Gaussian random variable with zero average and unit variance. In the GARCH volatility process,  $\sigma_t^2$  is defined by

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2)$$

where  $\alpha, \beta$  and  $\omega$  are the model parameters to be estimated.

To estimate the model parameters, we use a likelihood function (LF) of the GARCH model, defined by

$$\text{LF} = \prod_{t=1}^N \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{r_t^2}{2\sigma_t^2}\right). \quad (3)$$

The maximum likelihood estimation determines the model parameters by maximizing the LF. In the Bayesian estimation, the probability distribution of the parameters are constructed by using the LF under the Bayes theorem and then the parameters are obtained as expectation values over the probability distribution.

## III. ESTIMATION BY MACHINE LEARNING

In machine learning we determine the model parameter by minimizing a loss function defined by

Loss function =  $-\log(\text{LF})$

$$= \sum_{t=1}^N \frac{1}{2} \left( \log(\sigma_t^2) + \frac{r_t^2}{\sigma_t^2} + \log(2\pi) \right). \quad (4)$$

Minimizing the loss function leads to the same results with maximizing the likelihood function. To perform the machine learning, we use TensorFlow [19]. TensorFlow prepares several optimizers for minimization purpose. In this study we use the adaptive moment estimation (Adam) [20] which is known to show a better convergence property than other optimizers such as the gradient decent method. The Adam has several tunable parameters. We use default settings recommended in [20] except for the learning rate (LR). Varying LR, we perform the GARCH parameter estimation with the Adam optimizer and investigate its performance, convergence rate and so on.

IV. SIMULATION STUDY

First we generate artificial GARCH time series by eq.(2) with known parameters. Then we perform the parameter estimation by machine learning to the GARCH time series. The results will be compared with the known parameters. The computational environment of this study is summarized on Table 1.

Table 1: Computational environment.

CPU	Intel Core i7-8700 3.20GHz
Memory	16GB
OS	Windows 10
TensorFlow	Version 1.9.0
Python	Version 3.6.6

Fig.1 shows the GARCH time series generated with a set of input parameters,  $\alpha, \beta, \omega = (0.1, 0.8, 0.1)$ . The number of the data generated is 2000. We iterate the minimization process by the Adam optimizer 10000 times for various values of LR.

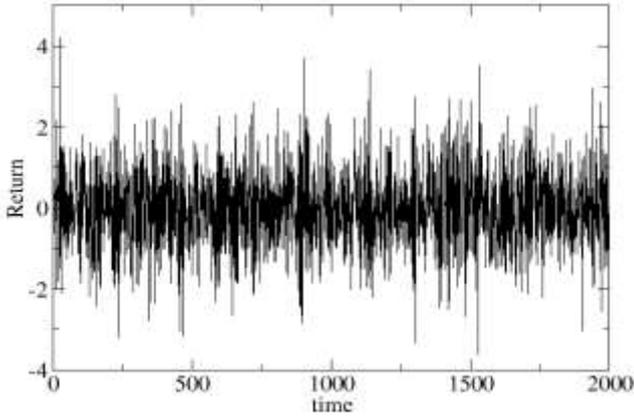


Fig.1 GARCH time series generated at  $\alpha, \beta, \omega = (0.1, 0.8, 0.1)$ .

Figs.2-4 show  $\alpha, \beta$  and  $\omega$  as a function of iteration, respectively. The initial values of parameters are set to  $\alpha, \beta, \omega = (0.3, 0.6, 0.05)$  and then the Adam optimizer is performed. For all the values of LR in  $[0.0001, 0.005]$ , the estimated values of  $\alpha, \beta$  and  $\omega$  seem to converge to the correct values. The precise values obtained are listed on Table 2. The values of  $\alpha, \beta, \omega$  estimated by the Bayesian method [12-17] applied to the same GARCH time series are also listed on Table 2. It is recognized that the values estimated by machine learning are consistent with those obtained by the Bayesian method.

It is found that the convergence speed to the correct values increases with increasing LR. However we also find that for values over LR=0.00547, the loss function by the Adam optimizer diverges. Thus there exists a maximum LR that preserves the convergence.

Figs.5-7 show trajectories of parameters from initial values to the converged values in  $\alpha - \beta, \beta - \omega$  and  $\alpha - \omega$  planes, respectively. We find that as the LR increases the trajectory starts to take a detour. It is understood that by taking a detour the Adam optimizer accelerates the convergence speed.

Table 2: Input parameters and estimated parameters by machine learning and Bayesian method.

	$\alpha$	$\beta$	$\omega$
Input parameters	0.1	0.80	0.1
Machine Learning	0.0967	0.802	0.0991
Bayesian method	0.103	0.779	0.115

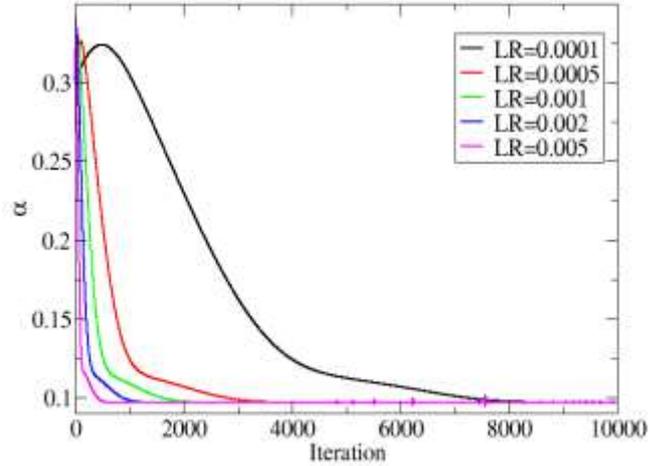


Fig.2 Parameter  $\alpha$  as a function of iteration for various LR.

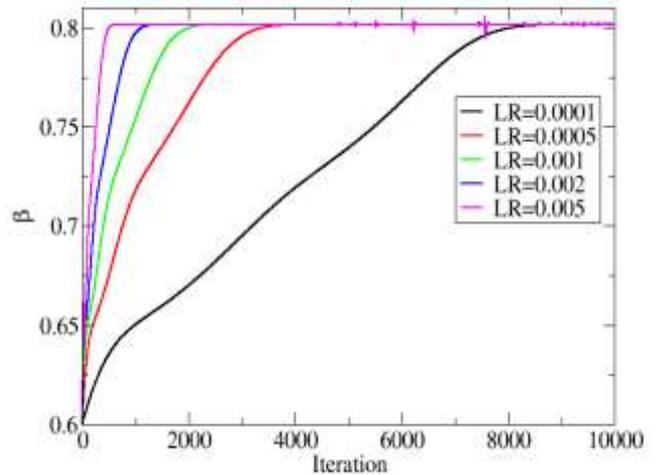


Fig.3 Parameter  $\beta$  as a function of iteration for various LR.

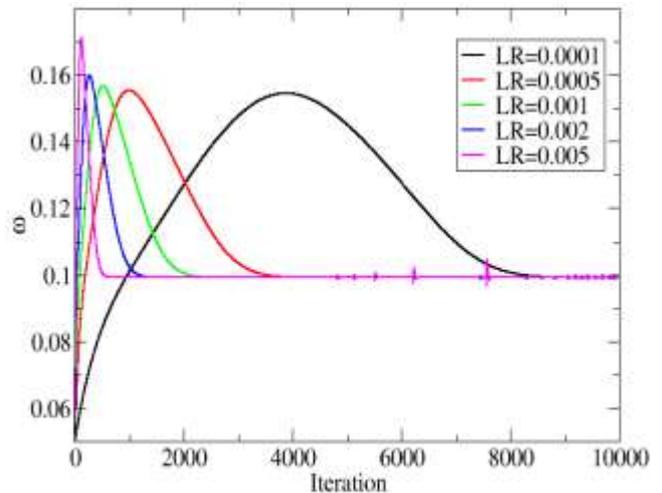


Fig.4 Parameter  $\omega$  as a function of iteration for various LR.

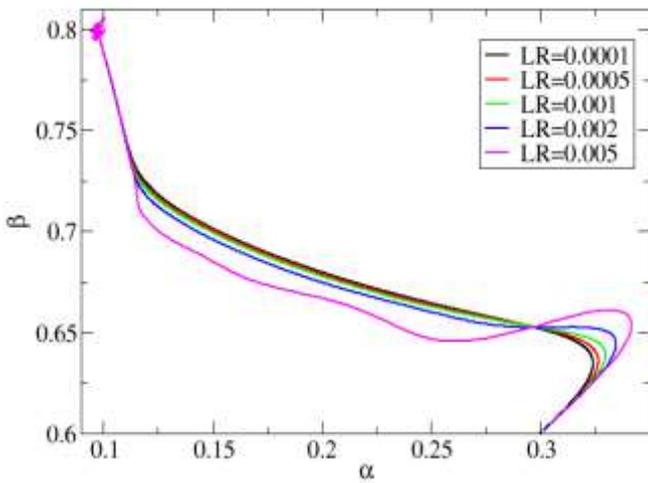


Fig.5 Parameter trajectory in  $\alpha - \beta$  plane.

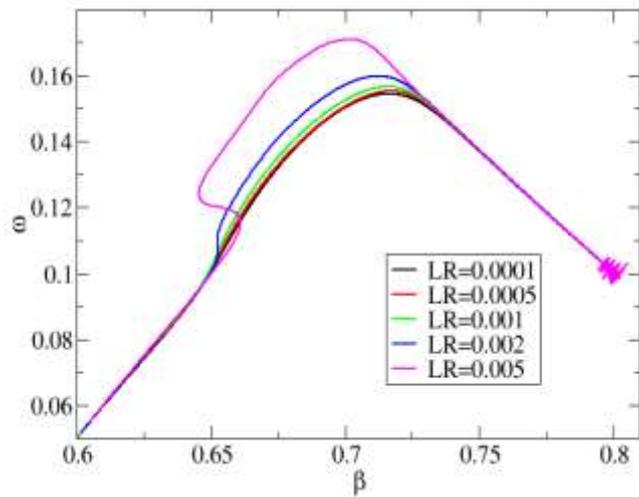


Fig. 6 Parameter trajectory in  $\beta - \omega$  plane.

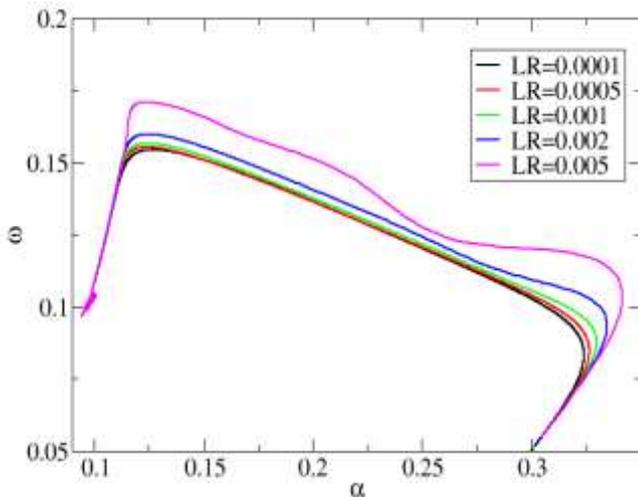


Fig.7 Parameter trajectory in  $\alpha - \omega$  plane.

In order to investigate the convergence rate, we calculate the error defined by  $error = Loss(iteration) - \min(Loss)$ . Fig.8 shows the error versus iteration for various LR. At sufficiently large iterations, the error seems to decrease exponentially. To measure the convergence rate, we fit the error by a function of  $a \exp(-\mu x)$  in a range of  $error=[0.0005,0.1]$ , where

$x$  stands for the iteration.  $a$  and  $\mu$  are the fitting parameters. We define  $\mu$  as a convergence rate. Fig.9 shows the convergence rate  $\mu$  as a function of LR. The convergence rate increases as LR increases, almost linearly up to around LR=0.005. For LR over 0.00547, the loss function by the Adam optimizer does not converge.

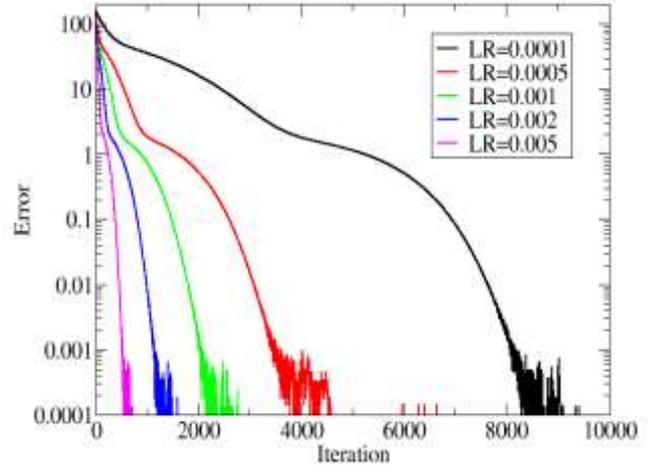


Fig.8 Error versus iteration for various LR.

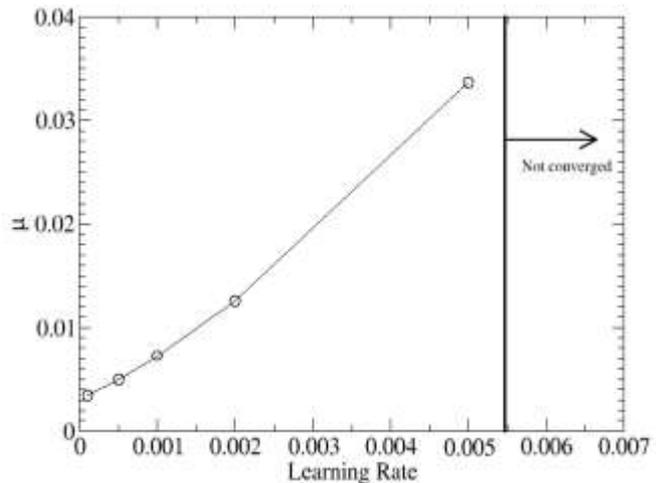


Fig.9 The convergence rate  $\mu$  as a function of LR.

## V. CONCLUSION

The GARCH parameter estimation is performed by machine learning. We apply the machine learning for the parameter estimation of the artificial GARCH time series generated with known parameters. By minimizing the loss function with the Adam optimizer of TensorFlow, we find that the model parameters of the GARCH model are determined correctly. The convergence property of the Adam optimizer is also investigated. We find that the convergence rate increases as the learning rate (LR) increases up to around LR=0.005. Over LR=0.00547, the Adam optimizer fails to minimize the loss function of the GARCH model.

This study focuses on a univariate GARCH model that has three model parameters. It would be interesting to further investigate multivariate GARCH models [21] with many model parameters that demand much computational effort and to elucidate whether machine learning can estimate parameters of the multivariate GARCH models.

## ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP18K01556.

## REFERENCES

- [1] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, vol.31, 1986, pp.307-327.
- [2] R.F. Engle, "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom inflation," *Econometrica*, vol.50, 1982, pp.987-1007.
- [3] T. Bollerslev, R.Y. Chou and K.F. Kroner, "ARCH modeling in finance," *Journal of Econometrics*, vol.52, 1992, pp.5-59.
- [4] R. Cont, "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues," *Quantitative Finance*, vol.1, 2001, pp.223-236.
- [5] D. Nelson, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, Vol.59, 1991, pp.347-370.
- [6] L. Glosten, R. Jagannathan, D. Runkle, "On the relation between the expected value and the volatility of the nominal excess on stocks," *J. Finance*, Vol.48, 1993, pp.1779-1801.
- [7] R.F. Engle and V. Ng, "Measuring and testing the impact of news on volatility," *Journal of Finance*, vol.48, 1993, pp.1749-1778.
- [8] E. Sentana, "Quadratic ARCH models," *Review of Economic Studies*, vol.62, 1995, pp.639-661.
- [9] T. Takaishi, Rational GARCH model: An empirical test for stock returns," *Physica A*, vol.473, 2017, pp.451-460.
- [10] T. Takaishi, "Statistical properties and multifractality of Bitcoin," *Physica A*, vol.506, 2018, pp.507-519."
- [11] L. Bauwens, M. Lubrano, "Bayesian inference on GARCH models using the Gibbs sampler," *Econometrics Journal*, vol.1, 1998, pp.23-46.
- [12] T. Takaishi, "Bayesian Estimation of GARCH model by Hybrid Monte Carlo," *Proceedings of the 9th Joint Conference on Information Sciences*, 2006, CIEF-214. Available: <https://doi.org/10.2991/jcis.2006.159>
- [13] T. Takaishi, "An Adaptive Markov Chain Monte Carlo Method for GARCH Model," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Complex Sciences*, vol.5, 2009, pp.1424-1434. Available: [https://doi.org/10.1007/978-3-642-02469-6\\_22](https://doi.org/10.1007/978-3-642-02469-6_22)
- [14] T. Takaishi, "Bayesian Estimation of GARCH Model with an Adaptive Proposal Density," *New Advances in Intelligent Decision Technologies, Studies in Computational Intelligence*, vol.199, 2009, pp.635-643. Available: [https://doi.org/10.1007/978-3-642-00909-9\\_61](https://doi.org/10.1007/978-3-642-00909-9_61)
- [15] T. Takaishi, "Bayesian Inference on QGARCH Model Using the Adaptive Construction Scheme," *Proceedings of 8th IEEE/ACIS International Conference on Computer and Information Science*, 2009, pp.525-529. Available: <https://doi.org/10.1109/ICIS.2009.173>
- [16] T. Takaishi, "Bayesian inference with an adaptive proposal density for GARCH models," *J. Phys.: Conf. Ser.*, vol.221, 2010, 012011.
- [17] T. Takaishi and T.T. Chen, "Bayesian Inference of the GARCH model with Rational Errors," *International Proceedings of Economics Development and Research*, vol.29, 2012, pp.303-307. Available: <http://www.ipedr.com/vol29/55-CEBMM2012-R00014.pdf>
- [18] T. Takaishi, "Markov Chain Monte Carlo versus Importance Sampling Bayesian Inference of the GARCH model," *Procedia Computer Science*, vol.22, 2013, pp.1056 - 1064.
- [19] TensorFlow <https://www.tensorflow.org/>
- [20] D.P. Kingma and J.L. Ba, "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION," arXiv:1412.6980. Available: <https://arxiv.org/pdf/1412.6980.pdf>
- [21] L. Bauwens, S. Laurent and J.V. Rombouts, "Multivariate GARCH models: a survey," *Journal of applied econometrics*, vol.21, 2006, pp.79-109.