Big data, big challenges: Risk management of financial market in the digital economy

Abstract

Purpose – The purpose of the research is to assess the risk of the financial market in the digital economy through the quantitative analysis model in the big data era. It's a big challenge for the government to carry out financial market risk management in the big data era.

Design/methodology/approach – In this study, a generalized autoregressive conditional heteroskedasticity-vector autoregression (GARCH-VaR) model is constructed to analyze the big data financial market in the digital economy. Additionally, the correlation test and stationarity test are carried out to construct the best fit model and get the corresponding VaR value.

Findings – Owing to the conditional heteroscedasticity, the index return series shows the leptokurtic and fat tail phenomenon. According to the AIC (Akaike Information Criterion), the fitting degree of the GARCH model is measured. The AIC value difference of the models under the three distributions is not obvious, and the differences between them can be ignored.

Originality/value – Using the GARCH-VaR model can better measure and predict the risk of the big data finance market and provide a reliable and quantitative basis for the current technology-driven regulation in the digital economy.

Keywords: Big data, digital economy, risk management, financial market, GARCH model, GARCH-VaR model.

Paper type Research paper

1. Introduction

The digital economy era of the domestic big data financial market started in 2017, and new scientific and technological terms such as artificial intelligence, blockchain, and big data have sprung up in front of the world. Unfortunately, in this year, economic chaos began to appear in the financial market, such as illegal ICO (Initial Coin Offering) etc. (Zyskind and Nathan, 2015; Shi *et al.*, 2020; Bo *et al.*, 2021; Yang *et al.*, 2021).

Although the process of financial innovation caused by science and technology is gradually advancing, economic management and legal protection have not kept pace with the times, and the scientific and technological innovation is often separated from the current supervision mechanism (Li *et al.*, 2020a; Liu *et al.*, 2020; Cai *et al.*, 2021).

As a result, the behavior of evading supervision in disguise often occurs, leading to regulatory arbitrage (Zhang *et al.*, 2020a; Chen and Liu, 2021). The above situation shows that China should build and optimize the financial regulatory mechanism and adapt to the financial innovation environment promoted by science and technology. Therefore, with the emergence of Artificial Intelligence (AI), blockchain, big data, and other high technologies in the digital economy, although the service link is simplified and the efficiency is improved, more complicated and rich risk situations appear one after another. Therefore, supervision departments should have the idea of "using science and technology to control science and technology" and build technology-driven risk control, which is a new topic for supervision departments to balance encouraging innovation and strengthening supervision.

Big data can expand the scope of financial market risk management (Cerchiello and Giudici, 2016; Sun *et al.,* 2019; Dong *et al.,* 2021; Zhang *et al.*, 2020b). Big data has the characteristics of large quantity, fast speed, and many kinds. In the era of the digital economy, the relationship between big data and the financial market is becoming closer and closer. In the financial market, big data records the information of individual and enterprise customers through text, video, and other storage media. Through big data, users' portraits can be outlined and clustered, which is conducive to accurate marketing by banks and other financial institutions. In the era of the digital economy, big data is the core asset of financial institutions (Pfeiffer and Review, 2019; Gong *et al.,* 2020). At the same time, this will bring a series of financial risk management problems. The data of traditional financial risk management is often scarce, so it is unable to deal with the risk in time and effectively. By means of big data information technology, we can build a digital financial risk management and control system, which will provide solutions for preventing financial risks.

With the arrival of the big data wave, the supervision of the financial market has also

brought changes. Use big data information technology such as data mining or neural networks to measure the financial market's risk. Monitor the relevant indicators of financial market risk, and realize the prediction and prevention of financial market risk. First, big data can provide technical means for financial market risk supervision. Big data can make financial supervision more credible and visual. Secondly, in the era of the digital economy, digital financial supervision is more accurate. The traditional financial supervision is often a unified supervision paradigm, which can not achieve personalized and differentiated supervision. Big data can make this beautiful vision a reality. The vigorous development of the financial market puts forward higher requirements for the ability of the Chinese government's regulatory authorities. The scale of financial supervision should be grasped in place. Government regulators should make rational use of financial risk monitoring tools such as big data, give full play to the information technology advantages of big data, and improve the government supervision system.

Previous literature mainly focused on risk prevention and diffusion. These studies can provide guidance for the business community and academia in some specific financial situations but cannot provide an effective reference in the era of the digital economy (Farag and Johan, 2021; Gravina and Lanzafame, 2021; Elheddad *et al*., 2021). In the aspect of risk early warning of financial technology, this paper thinks that it can be explored from the following three aspects: (1) In the area of the traditional digital financial risk, many scholars in the past focused on the characteristics of digital finance, but ignored the systematic reasons of the regulatory system. (2) The measurement of digital financial risk ignores the situation that the mode of business operation in China is contrary to the traditional financial theory, which will lead to differences in the measurement of digital financial risk. (3) In the era of the digital economy, the stability of the economy and media opinion have a great impact on the risk supervision of digital finance.

Therefore, the main contributions of this paper are as follows: This paper deeply analyzes the risk measurement indicators in the field of financial technology and uses the GARCH-VaR model to study the early warning and control of big data financial

technology risk. Through the value at risk calculated by the GARCH-VaR model, we can find the meaning behind the big data and provide technical support for intelligent supervision. There is still a big gap between traditional financial supervision and expectations. Through big data, blockchain, and other technical methods, China's regulatory system can be comprehensively upgraded. From a systematic perspective, this paper studies the risk prevention and early warning of digital finance and provides ideas for constructing and improving China's financial technology regulatory system.

The research aims to assess the risk of the financial market in the digital economy through the quantitative analysis model. First, a generalized autoregressive conditional heteroskedasticity-vector autoregression (GARCH-VaR) model is constructed to analyze the big data financial market in the digital economy. Additionally, the correlation test and stationarity test are carried out to construct the best fit model and get the corresponding VaR value. Based on the above discussions, this paper intends to discuss the following four issues from the theoretical and empirical levels:

i. By combing the risk types of China's digital financial market, this paper summarizes its risk characteristics.
ii. Examine the relationship between the digital financial market and government regulation.
iii. Calculate the VaR calculation and test of the digital financial market index.
iv. Optimize and improve the digital financial risk supervision standards and system.

The article is structured as follows: Section 2 provides the concepts such as big data financial market through an exhaustive literature review. Section 3 outlines the research methodology adopted for this article. Results and discussion have been presented in Section 4. Finally, implications to theory and practice of this article and key lessons learned have been presented along with the conclusion in Section 5.

2. Literature review

Under the background of discussing big data financial market supervision theory, many scholars still have blind spots in the research on risk management of big data financial market, which is reflected in the following aspects (Zhang *et al.*, 2019a). Firstly, the

discussion on the causes of financial technology risk only stays at the source, and also, there is a lack of research on supervision mechanisms and management systems. Secondly, more attention is paid to the risk research of financial products. Thirdly, it focuses on static and traditional research and lacks research on innovation theory and mechanism reform. Finally, most of the researchers focused on the control of management objectives and management mechanisms and often neglected the in-depth exploration of existing management innovation (Hu *et al.*, 2020a; Zhang *et al.*, 2019b).

The big data financial market is unpredictable and more complex than network risk and economic crisis. In fact, it also includes operational risk, liquidity risk, and market risk, which is similar to the traditional financial risk (Li *et al.*, 2020b; Hu *et al.*, 2020b; Zhang and Shen, 2021; Zhou et al., 2020). Moreover, some risks are unique, such as private security risks, technology risks, etc., which are more difficult to predict and avoid. The risk characteristics of the big data financial market are as follows:

Firstly, it has strong liquidity. The big data financial market is based on electronic communication, network engineering, and other increasingly prosperous technology that can be raised (Hu *et al.*, 2020c). It adopts various technologies to complement each other, thus carrying out various businesses and breaking the restrictions on space and time in the financial field (Jiang *et al.*, 2018; Hu *et al.*, 2020d). It makes financial communication more convenient using big data technology and improves the efficiency of capital allocation and the flow rate. Of course, it may also have a huge impact on the development of the financial sector due to the negative effects of big data in the digital economy.

Secondly, there is a wide range of risks. As an emerging technology in the new era, the big data financial market relies on mobile communication technology for development, with the characteristics of openness, universality, and wide range. Although it has greatly promoted the smooth communication of the financial industry in the world, the consequences will be multiplied. The risk correlation between regions will be stronger, and the scope of risk will be wider.

Thirdly, it has high concealment. In the related operations of big data technology for the financial industry, most of the financial business through the website or app

registration and other ways (Niu *et al.*, 2020; Hua *et al.*, 2021; Lv *et al.*, 2021a). Although the online operation greatly reduces the labor cost and the cost of financial transactions, the financial risks become more unpredictable and controllable (He *et al.*, 2020a; Wang *et al.*, 2021a; Jing *et al.*, 2021; Lv *et al.*, 2021b). Compared with the traditional financial industry, the big data financial industry breaks through the multiple constraints of time and space in the traditional financial industry, making different kinds of businesses penetrate and connect with each other (Zuo *et al.*, 2015; He *et al.*, 2020b; Liu *et al.*, 2021; Ma *et al.*, 2021). On the other hand, it brings greater challenges to the risk supervision of big data.

Fourthly, it has complex characteristics as the big data financial market provides open financial services for customers, customers' financial transactions tend to have a high dependence on Internet security technology (Veselovsky *et al.*, 2018; Bhimani and Willcocks, 2014; Ni *et al.*, 2021; Sun and Lv, 2021; Wang *et al.*, 2021b). The application of internet technology in the financial industry has many complex nodes, such as ineffective supervision and low technical level. Moreover, with the continuous development of big data financial technology, it gradually integrates into a variety of businesses, such as banking, fund, insurance, etc., and its application scope is wider and wider (Brynjolfsson *et al.*, 2003; Curran, 2020; Akter and Wamba, 2016; Ouyang *et al.*, 2021; Wang *et al.*, 2021c). Therefore, the complexity of the big data financial market is further enhanced.

3. Methodology

Because of the uncertainty in the financial market, the cost of big data financial technology products or services will fluctuate greatly. Most scholars used the GARCH (Generalized Auto Regressive Conditional Heteroskedasticity) model to analyze the volatility of the financial market price index series to determine the law of volatility and the difference inaccuracy (Zikmund, 1994). Furthermore, the AIC (Akaike Information Criterion) criterion is used to simulate the conditional variance, which is more suitable for reality. And the VaR (Value at Risk) method is mainly used to measure the big data financial technology risk. The paper calculates the conditional variance mean and quantile of the return series by the GARCH model and gets the value

at risk of the big data financial market.

3.1 Applicability of GARCH-VaR model

GARCH-VaR model absorbs the advantages of both the VaR method and GARCH model (McAleer and Da Veiga, 2008). It has a wide range of measurements because the GARCH model can accurately explain the financial data with the leptokurtic and fat tail phenomenon. From the perspective of risk supervision, the distribution of leptokurtic and fat tails is due to the clustering and concentration characteristics of price fluctuations. The peak of random variables appears in the average range, and the events with higher risk are small probability events. The impact of price fluctuation is mainly caused by the variance of time series and may also be affected by various objective environments, such as policies, natural disasters, wars, etc.

3.2 Data selection

At present, there are both traditional Internet financial models and emerging big data financial technology models in the financial market. Therefore, this paper selects the wind Internet financial index (code 884136) to measure the market risk faced by financial technology-listed companies. The sample involves third-party payment, e-commerce banking, and other industries, which are highly representative and reliable and can reflect the overall situation of the big data financial market (Li *et al.*, 2021).

The data sample of this paper is the daily closing price of the science and technology index from January 2014 to December 2020. China's big data financial market has entered a period of stable development after 2014. In this paper, wind Internet financial index is selected to study, which can reduce the interference of special data to the results of this study and make the results of model fitting more accurate.

When describing the rate of return, the sample data is logarithmically processed. The main reason is that the similarity level of continuous compound interest and logarithmic rate of return in meaning is high. Logarithmic processing of sample data to represent the rate of return can eliminate the non-stationarity of the series and further improve the fitting effect of the model (Lv and Liu, 2021; Ma *et al.*, 2021). The logarithmic first-order difference form of the daily yield of big data financial technology index is as follows:

$$r_t = \ln(\frac{sp_t}{sp_{t-1}}) = \ln Sp_t - \ln Sp_{t-1} \quad (1)$$

which Sp_t represents the daily closing price of the big data financial market index and r_t represents the yield of the financial market index on the $t-th$ trading day.

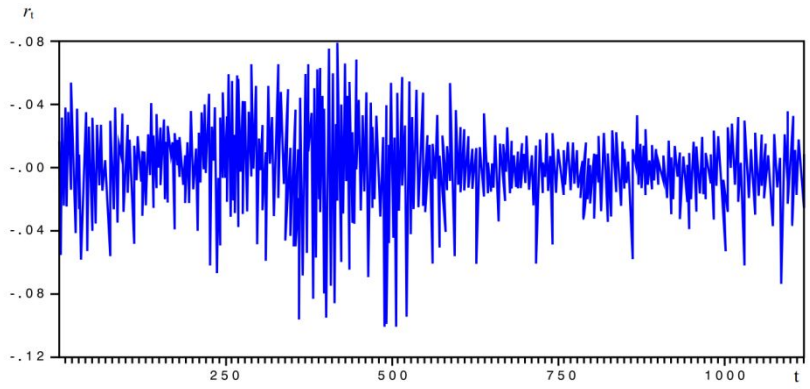

Figure 1. Rate of return r_t series volatility chart

Figure 1 is the rate of return r_t series volatility chart. The fluctuation of the exponential, logarithmic rate of return is random, and the whole fluctuation chart appears obvious clustering phenomenon. From the 700th observation to the 900th observation, the fluctuation is small. From the 250th observation to the 550th observation, the observation fluctuates greatly, and there are significant positive returns and negative returns at the same time. This phenomenon is called volatility autocorrelation. The size of volatility is often related to its income. The greater the volatility, the higher the income, and the smaller the volatility, the smaller the income.

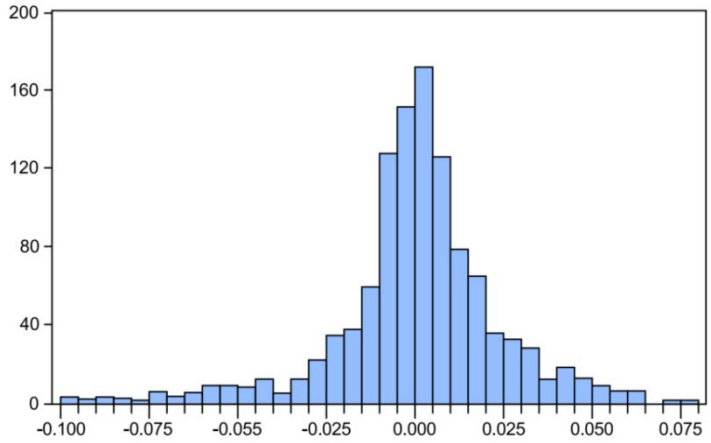

Figure 2. Statistical analysis chart of r_t

From the data shown in Figure 2, the mean value of the return series is 0.000808, the standard deviation and skewness are 0.023133 and -0.6949, respectively. Considering the morphological characteristics of data skewness distribution, if the skewness coefficient is less than 0, it is proved that the data distribution of the mean value of the return series has a long tail on the left. Kurtosis can be used to measure the change of the distribution situation of the whole value. From the data shown in the figure, the kurtosis of the series value is 5.755980, which proves that the yield series has the characteristics of a thick tail and cusp. If the Jarque-Bera statistic is 443.8016 and the p-value is 0.00000, the hypothesis that the return series r_t follows normal distribution is rejected.

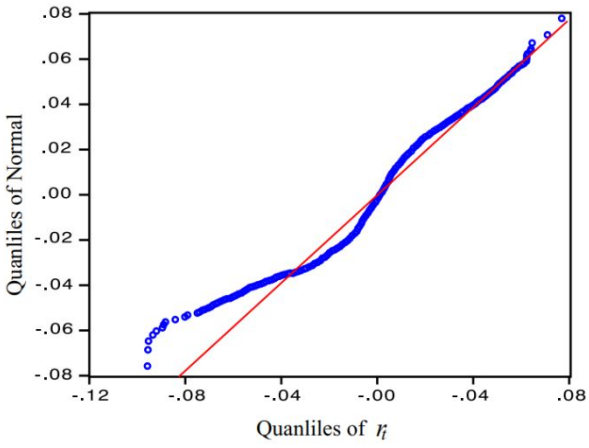

Figure 3. QQ Chart of r_t

The line shown in Figure 3 represents the specific distribution of the return sequence

under the normal distribution. And the actual distribution r_t is represented by a curve scatter graph. There is a big deviation between the curve scatter graph and the straight line near the horizontal axis from the change of the curve scatter graph and the change of straight line in the graph. However, the curve scatter graph in the middle is distributed close to the straight line, and the curve scatters graph, and the straight line is close to each other, far away from the horizontal axis, which proves that the yield sequence r_t has the feature of fat tail distribution.

4. Results and discussion

4.1 Stationarity and unit root test

Whether the time series is stable or not is crucial to the accuracy of the whole analysis results. When two non-stationary series have the trend of phase change, the statistical information may be misguided. Therefore, in the time series test, we must test the stationarity of the series to avoid the occurrence of wrong research conclusions. The results of the stationary test are shown in Table 1.

Table 1. ADF stationarity test

	ADF test	1% level test threshold	5% level test threshold	10% level test threshold
t-statistic	-28.8358	-3.3248	-2.7528	-2.4571

In the process of testing whether the time series is stable or not, this paper uses the unit root as an indicator to test. According to the stationarity test results, there is stationarity in the 10% confidence interval, the tested coefficient is -3.3248, and the stationarity coefficients in the 5% and 1% confidence intervals are -2.7528 and -2.4571, respectively. Considering that the exponential series has a unit root, the original hypothesis is not tenable. There is no unit root in the sequence. At the same time, it also proves that r_t sequence belongs to the stationary sequence.

4.2 ARCH effect test

After the stationarity test and unit root test, we need to test whether the whole sequence has an ARCH effect and establish a random walk model as follows:

$$\ln Sp = \mu + \rho \times \ln Sp_{t-1} + u_t \qquad (2)$$

This paper uses the least square estimation method to test the effect of residual sequence in the testing process, and the results are shown in Table 2.

Table 2. Estimation results of random walk model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNSP(-1)	0.997782	0.001287	714.2810	0.0000
R-squared	0.998811	Mean dependent VaR		8.322135
Adjusted R-squared	0.998809	S.D.dependent VaR		0.482140
S.E. of regression	0.023078	Akaike info criterion		-4.575103
Sum squared resid	0.593850	Schwarz criterion		-4.577124
Log likehood	2636.571	Hannan-Quinn criterion		-4.582708
F-statistic	507531.2	Durbin-Watson		1.717114
Prob(F-statistic)	0.0000			

We can get formula (3) from Table 2.

$$\ln Sp = 0.027853 + 0.996793 \ln Sp_{t-1} + \hat{u}_t \tag{3}$$

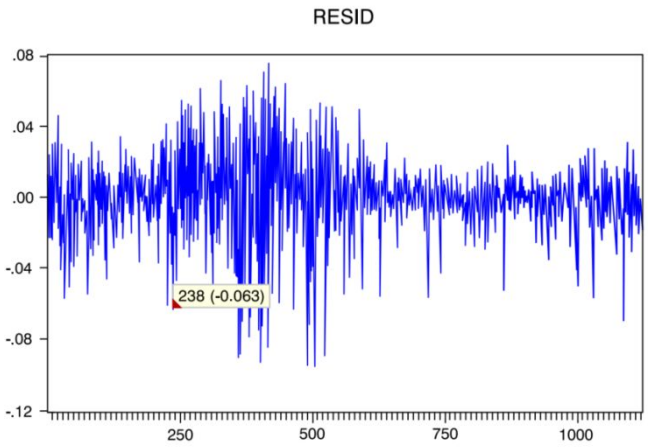

Figure 4. Residual of OLS regression equation

According to Figure 4, when the *F* statistic is large enough, the significance of the calculation results is obvious enough. $R^2 = 0.9988$ shows that the fitting effect of the model is good. The coefficient of the explanatory variable $\ln Sp(-1)$ is estimated to be 0.9978, indicating that it is significant. After observing the residual diagram of the

regression equation, we find that the {SP} fluctuation "cluster" phenomenon is small between 700 ~ 900 and large between 400 ~ 500, which proves that the sequence may have conditional heteroscedasticity.

From the residual square correlation graph and the specific data, we know that the partial autocorrelation coefficient and autocorrelation coefficient of the residual sequence are mostly above the 95% confidence level, and the Q statistic is also significant. Table 3 shows that the big data financial market index series has an ARCH effect.

Table 3. Test of ARCH effect of residual sequence

F-statistic	95.80214	Prob.F(3,111)	0.0000
Obs*R-squared	229.2415	Prob.Chi-Square(3)	0.0000

From the test results given in Table 3 and the calculation results of F statistics, we can know that the corresponding value of F statistics is 95.80214, and the corresponding p-value is zero, which proves that all residual sequences in the regression equation are significant. The statistical value of the ARCH effect T× R_2 is 229.2415, and its probability p-value is 0, so the hypothesis that the residual does not conform to arch characteristics is not tenable. The results show that the ARCH effect exists in both the big data financial market index and yield index. Therefore, it is feasible to study the volatility of yield in China's big data financial market by using the GARCH model.

4.3 Establishment of GARCH model

This paper used the normal distribution of random error term to describe its distribution characteristics more accurately. When the sequence satisfies the normal distribution, the following equation can be obtained.

Mean value equation: $r_t = -4.25\times10^{-5} + 0.1012\hat{\sigma}_t^2 + \mu_t$ (4)

Variance equation: $\hat{\sigma}_t^2 = 2.94\times10^{-6} + 0.0504\hat{u}_{t-1}^2 + 0.94104\hat{\sigma}_{t-1}^2$ (5)

When the coefficients of three parameters in conditional variance are all positive,

each parameter is highly significant. The coefficient $\hat{\sigma}_t^2$ is 0.1012, which proves that the yield increases by 0.1012% when the expected risk increases by one unit. We can know that the final value of the variance coefficient of the lag condition is the final value of $(\hat{\alpha}+\hat{\beta})$ is 0.992, which meets the standard of parameter value and not greater than the constraint condition. The closer the final value is to 1, the longer the influence persistence of conditional variance is. Moreover, it is proved that stationary fitting data of the GARCH (1,1) model is better. Furthermore, this paper tests conditional heteroscedasticity. When p = 3 (lag coefficient), the statistics of the residual series are shown in Table 4.

Table 4. Test results of ARCH LM

F-statistic	0.513133	Prob.F(3,111)	0.5615
Obs*R-squared	1.544665	Prob.Chi-Square(3)	0.5616

The probability value of the big data financial market index return series r_t is 0.5615, which does not conflict with the previous hypothesis. Therefore, we can conclude that the residual series has no conditional heteroscedasticity effect, and we can use GARCH (1,1) model to eliminate the ARCH effect.

AC (Autocorrelation Parameter) and PAC (Partial Autocorrelation Parameter) tend to 0, and the statistic Q tends to 1. The height of P (residual sequence probability) is not significant, which indicates that the ARCH effect has been eliminated by GARCH (1,1) model. The parameters of the first-order GARCH model of GED and student's t distribution are calculated in the same way.

Table 5. GARCH (1,1) estimation under three distributions

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Distribution type	ω	α	β
Normal distribution	2.93×10^{-6} (0.00**)	0.050 (0.00**)	0.931 (0.00**)
Student's t	1.61×10^{-6}	0.065	0.912

	(0.1644)	(0.00**)	(0.00**)
GED	2.44×10^{-6}	0.053	0.941
	0.0231	(0.00**)	(0.00**)

Note: * * means significant at 99% confidence level

Table 5 is GARCH (1,1) estimation under three distributions. The conditional variance coefficient (CV) is calculated using the parameters of normal distribution type, etc., and the results are all less than 1, indicating that the variance stability is significant in these three distribution types. If $\alpha>0$, it means that the GARCH effect exists in the big data financial market. If $\beta>0.9$, it indicates that non-real-time information can still affect market volatility to a great extent. It can be seen from Table 5 that these parameters have little difference in different distribution types, which indicates that the return rate of the big data financial technology index is not normally distributed.

Table 6. GARCH (1,1) - M estimation for three distributions

Distribution type	ω	α	β_1	β_2
Normal distribution	2.97×10^{-6} (0.00**)	0.0533 (0.00**)	0.9376 (0.00**)	0.2402 (0.007*)
Student's t	2.05×10^{-6} (0.124)	0.072 (0.00**)	0.925 (0.00**)	0.279 (0.0001**)
GED	$2.77\times10_{-6}$ (0.0237*)	0.062 (0.00**)	0.932 (0.00**)	0.325 (0.00**)

Note: * means significant at 95% confidence level

** means significant at 99% confidence level

Table 6 is GARCH (1,1)-M estimation for three distributions. The standard deviation coefficient $\beta_2>0$, which indicates that there is a positive risk premium. And the sample yield series is positively correlated with the market volatility. Currently, most consumers in the market are risk-averse consumers, requiring high risk and high return. When $\alpha+\beta_1<1$, GARCH (1,1) model is in a stable state and has significant valuation, which shows that the previous price fluctuation can greatly affect the current price fluctuation, and the past market risk can determine the current market risk.

Table 7. EGARCH (1,1) estimation under three distributions

Distribution type	ω	α	γ	β
Normal distribution	-0.156×10^{-6} (0.00**)	0.134 (0.00**)	-0.025 (0.0068*)	0.994 (0.00**)

Student's t	-0.142×10^{-6} (0.00**)	0.158 (0.00**)	-0.0178 (0.1777)	0.996 (0.00**)
GED	-0.155×10^{-6} (0.00**)	0.137 (0.00**)	-0.0234 (0.144)	0.995 (0.00**)

Note: * means significant at 95% confidence level

** means significant at 99% confidence level

Table 7 is EGARCH (1,1) estimation under three distributions. Coefficient γ represents the asymmetric effect. In normal distribution type, γ=-0.025. In student's t distribution, γ=-0.0178. In GED distribution, γ=-0.0234. The normal distribution of value γ has 95% significance, while the other two distributions have no significance. In a normal distribution, α= 0.134 and γ=-0.025. When $u_{t-1}-1>0$, the conditional variance will be impacted by 0.109 times of information shock. When $u_{t-1}-1<0$, the conditional variance will be impacted by 0.159 times of information shock. Referring to the theory of negative leverage effect, the negative impact will have a stronger impact on the next cycle than the positive impact of the same strength.

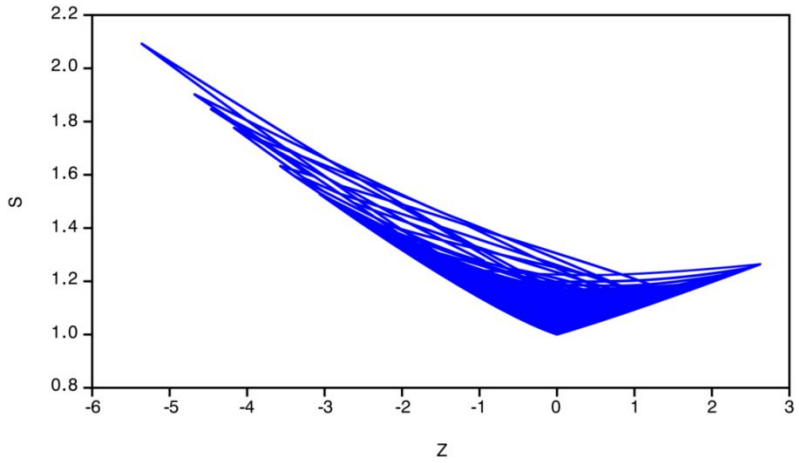

Figure 5. Asymmetric information shock curve

As can be seen from Figure 5, the information shock value shows a negative shock. When $u_{t-1}-1<0$, the impact curve is steeper. When $u_{t-1}-1>0$, it is in the frontal impact, and the curve is slower. The emergence of bad information will lead to stronger market volatility, and a negative impact has a greater impact on volatility.

T-ARCH (1.1) model can be used to verify the asymmetry of price fluctuation. The asymmetric term is negative in student's t distribution. There was no significant

difference among the three distributions. Therefore, the model has no asymmetric effect in this yield series. Table 8 is T-ARCH (1,1) estimation under three distributions

Table 8. T-ARCH (1,1) estimation under three distributions

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1} + \beta_1\sigma_{t-1}^2$$

Distribution type	ω	α	γ	β
Normal distribution	2.93×10^{-6} (0.00**)	0.055 (0.00**)	0.006 (0.611)	0.9501 (0.00**)
Student's t	1.62×10^{-6} (0.2155)	0.079 (0.00**)	-0.0012 (0.954)	0.936 (0.00**)
GED	2.43×10^{-6} (0.00**)	0.068 (0.003*)	0.002 (0.926)	0.9325 (0.00**)

Note: * means significant at 95% confidence level

** means significant at 99% confidence level

This paper proposes AIC can get the best fitting GARCH model while analyzing the equation parameters of the GARCH model. The basic theoretical concept is entropy, which can effectively evaluate the volatility of the GARCH model fitting time series and judge its excellent degree.

$$\text{Akaike information criterion(AIC)} = \ln(\sigma^2) + \frac{2K}{T} \quad (6)$$

In which T is the sample size, $\ln(\sigma^2)$ is the likelihood function, and K is the total number of parameters to be estimated. The smaller the AIC value, the better the fitting degree of the model.

Table 9. AIC of GARCH model

Model	Normal distribution	Student's t	GED
GARCH(1,1)	-5.113	-5.187	-5.174
GARCH(1,1)-M	-5.129	-5.202	-5.188
EGARCH(1,1)	-5.128	-5.203	-5.180
TARCH(1,1)	-5.112	-5.187	-5.172

Table 9 is the AIC of the GARCH model. The model of student's t and GED performs better than a normal distribution. The AIC value of the GARCH (1,1) - M

model in student's t distribution type is the smallest. The parameter ω is not significant, and the difference between the estimated values of parameters of all models in several distribution types is less than 0.09. It is necessary to combine the modeling complexity and modeling efficiency to select the best model for research.

4.4 Calculation and test of VaR

The calculation principle of VaR value is based on GARCH (1,1) model, which introduces the variance σ_t^2 that can be iteratively calculated on day t into the VaR formula. Therefore, the VaR value on t-th day can be obtained according to the calculation on t-1-th day. R^* represents the minimum rate of return, and α represents the significance level.

$$R^* = \mu - Z_1 - \alpha\sigma_t\sqrt{\Delta t} \tag{7}$$

$$VaR = W_{t-1}Z_1 - \alpha\sigma_t\sqrt{\Delta t} \tag{8}$$

The conditional variance in the GARCH model is normalized to get the conditional standard deviation, which is expressed by σ_t. μ represents the mean value, $Z_{1-\alpha}$ represents the quantile value, and Δt represents the unit within one day. The VaR value of each closing day can be obtained. At the same time, this paper calculates the VaR value under the condition of 90%, 95%, and 99% significance. Table 10 is descriptive statistics of VaR.

Table 10. Descriptive statistics of VaR

	90%	95%	99%
Mean value	0.488378	0.580342	0.764272
Median	0.093122	0.111717	0.146177
Maximum	4.524878	5.375321	7.076900
Minimum value	0.005344	0.006352	0.008366
Standard deviation	0.841726	1.000000	1.316216
JB statistic	2260.880	2260.880	2260.880
ProbabilityP	0.000000	0.000000	0.000000
Gross value	547.2374	650.0538	855.6861

Sum squares of total deviations	785.5042	1226.000	1944.225

According to the data in the Table, the order of VaR values under different confidence conditions is 99% > 95% > 90%. VaR (90%) is 0.488378, VAR (95%) is 0.580342, and VaR (99%) is 0.764272. The confidence level is positively correlated with the corresponding value at risk statistics. When the confidence level is 99%, VaR has the largest difference.

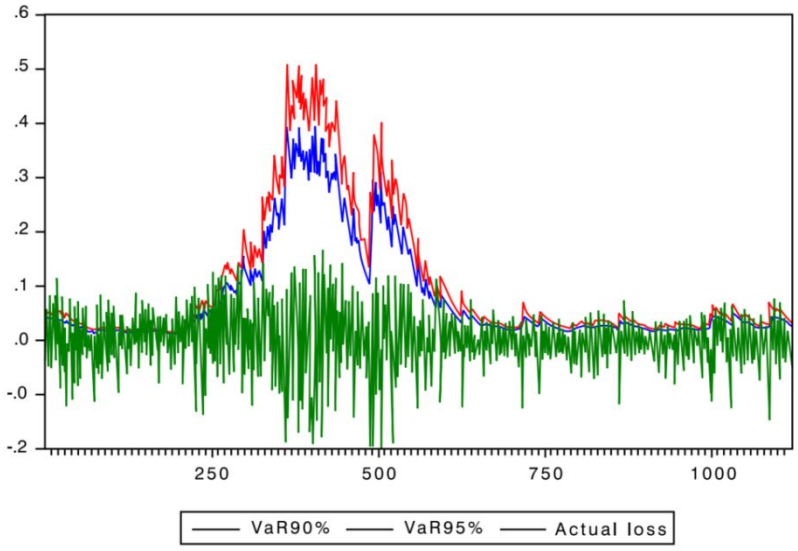

Figure 6. comparison of value at risk and actual loss

Figure 6 shows that when the sample is in the range of 300-550, VaR fluctuates greatly. It indicates that the volatility of the big data financial market index has increased significantly from 2014 to 2016, and the maximum risk appears when the sample value is 400. Especially in 2015, VAR increased rapidly in the first half of the year and reached the peak in the second half of the year. Now it is known that the stock market was turbulent at the beginning of 2015, and the sharp drop of stock price obviously affected the fluctuation of the big data financial market to a great extent.

In order to measure the accuracy of the result of VaR, the paper tests the value of VaR under various confidence levels and the actual loss and judge by the number of days of operation failure and the failure rate. It can be seen from the figure that the estimated result of VaR includes the actual loss and its degree.

Given confidence level α, the estimated failure frequency P^* is $1 - \alpha$. The actual failure probability is calculated by P. T is the total number of days of sample selection, W is the number of days of failure from real measurement.

Table 11. Failure days and failure probability under different confidence levels

VaR confidence level	99%	95%	90%
Failure days	62	72	87
Failure probability	5.2%	6.2%	7.8%

Table 11 is Under the confidence levels of 90%, 99%, and 95%, the failure day is 87, 62, and 72, respectively, and the failure probability is 7.8%, 5.2%, and 6.2%, respectively. Therefore, it can be judged that the calculation results of models with a 99% confidence level or less generally include the actual loss. Overall, the difference of failure probability among the three VaR confidence levels is not obvious. It can be seen from the failure days is significantly highlighted in 2014, indicating that the information transmission lags at the beginning of the development of the big data financial market. With the passage of time, the volatility of yield tends to be stable, and the number of failure days also decreases. VaR can better predict the loss. In addition, as far as the confidence level is concerned, the VaR of 99% confidence level is better. It means that under the 99% confidence level, regulators or investors can bear the larger loss preparation.

5. Conclusion

5.1 Implications to theory and practice

This paper takes the big data financial market as the research object and selects the daily closing price of the big data financial index as the sample data. Then the correlation test, stationarity test, and ARCH effect test are carried out to construct the best fit GARCH model and get the corresponding VaR value. First, the index return series shows the leptokurtic and fat tail phenomenon because of the conditional heteroscedasticity. Second, according to the AIC criterion, the fitting degree of the

GARCH model is measured. The AIC value difference of the models under the three distributions is not obvious, and the differences between them can be ignored. Third, the value at risk and the actual loss changed significantly from 2014 to 2016, with the peak value in the second half of 2015. This shows that the big data financial index entered a stage of rapid change in the first half of 2015, and its change trend is very large, so it has potential risks.

With the rapid development of financial science and technology, relying on science and technology to promote regulatory innovation and enhance Internet financial risk prevention, avoidance, and control is the key task and entry point for regulators to balance financial innovation and financial supervision. With the help of technology-driven thinking to guide supervision and management, the GARCH-VaR risk measurement model is incorporated into the big data risk monitoring system to help financial regulatory authorities carry out risk identification, identification, early warning, investigation, and other work. In addition, risk warning services can be provided to the public.

5.2 Key lessons learned

At present, the financial supervision system must keep up with the development of the times so that the financial risk and supporting the innovation and development of the real economy are in a state of balance (Wu *et al.*, 2020; Xu *et al.*, 2020; Xue *et al.*, 2020a; Xue *et al.*, 2020b; Zhang *et al.*, 2021). Therefore, regulators must implement long-term and targeted policies to avoid sudden strong intervention and stabilize market and policy expectations. In addition, to promote the continuous development of supervision and management technology, the exhibition is also the core content of macro-prudential management. To improve the professionalism, real-time, and penetration of supervision and management technology, we should build a supervision and management technology system relying on AI, digital technology, etc. For example, implementing a regulatory sandbox mechanism enables digital finance to have a corresponding safe range in developing the real economy and implementing incentive-compatible supervision and management policies. At the same time, we must also pay attention to the supervision of the whole process, strengthen the transmission efficiency

of regulatory information and data, and take the initiative to make digital finance form a strong new driving force for innovation and development. For example, the big data financial index can only estimate its market risk but cannot play a significant role in the process of risk estimation of Internet financial diversification and transitivity. Therefore, we must combine all kinds of measurement methods and supervision and management methods with playing a role.

5.3 Future research recommendations

In the aspect of big data financial risk supervision, this paper puts forward the following prospects: first, in the aspect of the financial technology business, we should rely on various capital and liquidity supervision indicators of financial institutions to improve the coverage of supervision and management in the field of the financial technology business, as well as cross-sectoral coordination work. At the same time, we should strengthen the risk management mechanism of individual institutions and introduce it into the unified macro framework to minimize the possibility of negative impact. Second, we should improve the information disclosure mechanism and cash custody mechanism of the financial technology business, enhance the openness of business practice, enhance investors' belief, and avoid the general panic in the industry caused by the pressure stage. Third, we must pay attention to the synchronization of strategy, the resonance effect caused by the algorithm, and implement the periodic counter adjustment. In addition, the regulatory authorities must also take the initiative to carry out supervision through regulatory technology and deeply combine the new technology and regulatory system to improve financial supervision efficiency. Fourth, in the aspect of digital currency, we must attach great importance to decentralization to avoid illegal financing and market manipulation; And for the design of legal digital currency, we must adhere to the principle of centralization and make efficient use of macro-control related functions.

References

Akter, S. and Wamba, S.F. (2016), "Big data analytics in E-commerce: a systematic review and agenda for future research", *Electronic Markets*, Vol. 26 No. 2, pp. 173-194.

Bhimani, A. and Willcocks, L. (2014), "Digitisation,'Big Data' and the transformation of

accounting information", *Accounting and Business Research*, Vol. 44 No. 4, pp. 469-490.

Bo, W., Fang, Z.B., Wei, L.X., Cheng, Z.F. and Hua, Z.X. (2021), "Malicious URLs detection based on a novel optimization algorithm", *IEICE TRANSACTIONS on Information and Systems*, Vol. 104 No. 4, pp. 513-516.

Brynjolfsson, E., Hu, Y. and Smith, M.D. (2003), "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers", *Management Science*, Vol. 49 No. 11, pp. 1580-1596.

Cai, K., Chen, H., Ai, W., Miao, X., Lin, Q. and Feng, Q. (2021), "Feedback convolutional network for intelligent data fusion based on near-infrared collaborative IoT technology", *IEEE Transactions on Industrial Informatics*, doi:10.1109/TII.2021.3076513.

Cerchiello, P. and Giudici, P. (2016), "Big data analysis for financial risk management", *Journal of Big Data*, Vol. 3 No. 1, p. 18.

Chen, J.J. and Liu, M.F. (2021), "Does the internet expand the educational gap among different social classes? The protective role of future orientation", *Frontiers in Psychology*, Vol. 12, p. 647351.

Curran, D. (2020), "Connecting risk: systemic risk from finance to the digital", *Economy and Society*, Vol. 49 No. 2, pp. 239-264.

Dong, S., Zhu, H., Zhong, S., Shi, K. and Liu, Y. (2021), "New study on fixed-time synchronization control of delayed inertial memristive neural networks", *Applied Mathematics and Computation*, Vol. 399, p. 126035.

Elheddad, M., Benjasak, C., Deljavan, R., Alharthi, M. and Almabrok, J.M. (2021), "The effect of the Fourth Industrial Revolution on the environment: The relationship between electronic finance and pollution in OECD countries", *Technological Forecasting and Social Change*, Vol. 163, p. 120485.

Farag, H. and Johan, S. (2021), "How alternative finance informs central themes in corporate finance", *Journal of Corporate Finance*, Vol. 67, p. 101879.

Gong, C., Hu, Y., Gao, J., Wang, Y. and Yan, L. (2020), "An improved delay-suppressed sliding-mode observer for sensorless vector-controlled PMSM", *IEEE Transactions on Industrial Electronics*, Vol. 67 No. 7, pp. 5913-5923.

Gravina, A.F. and Lanzafame, M. (2021), "Finance, globalization, technology and inequality: Do nonlinearities matter?", *Economic Modelling*, Vol. 96, pp. 96-110.

He, S., Guo, F., Zou, Q. and Hui, D. (2020a), "MRMD2.0: a python tool for machine learning with feature ranking and reduction", *Current Bioinformatics*, Vol. 15 No. 10, pp. 1213-1221.

He, S., Guo, F., Zou, Q. and HuiDing. (2020b), "MRMD2.0: a python tool for machine learning with feature ranking and reduction", *Current Bioinformatics*, Vol. 15 No. 10, pp. 1213-1221.

Hu, J., Wang, M., Zhao, C., Pan, Q. and Du, C. (2020a), "Formation control and collision avoidance for multi-UAV systems based on Voronoi partition", *Science China*

Technological Sciences, Vol. 63 No. 1, pp. 65-72.

Hu, J., Zhang, H., Li, Z., Zhao, C., Xu, Z. and Pan, Q. (2020b), "Object traversing by monocular UAV in outdoor environment", *Asian Journal of Control*, available at: https://doi.org/10.1002/asjc.2415.

Hu, J., Zhang, H., Liu, L., Zhu, X., Zhao, C. and Pan, Q. (2020c), "Convergent multiagent formation control with collision avoidance", *IEEE Transactions on Robotics*, Vol 36 No. 6, pp. 1805-1818.

Hu, J., Zhang, H., Li, Z., Zhao, C., Xu, Z. and Pan, Q. (2020d), "Object traversing by monocular UAV in outdoor environment", *Asian Journal of Control*. doi: 10.1002/asjc.2415.

Hua, L., Zhu, H., Shi, K., Zhong, S., Tang, Y. and Liu, Y. (2021), "Novel finite-time reliable control design for memristor-based inertial neural networks with mixed time-varying delays", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 68 No. 4, pp. 1599-1609.

Jiang, Q., Shao, F., Lin, W., Gu, K., Jiang, G. and Sun, H. (2018), "Optimizing multistage discriminative dictionaries for blind image quality assessment", *IEEE Transactions on Multimedia*, Vol. 20 No. 8, pp. 2035-2048.

Jing, L., Pan, Y., Wang, T., Qu, R. and Cheng, P.T. (2021), "Transient analysis and verification of a magnetic gear integrated permanent magnet brushless machine with Halbach arrays", IEEE Journal of Emerging and Selected Topics in Power Electronics. doi: 10.1109/JESTPE.2021.3057665.

Li, A., Spano, D., Krivochiza, J., Domouchtsidis, S., Tsinos, C.G., Masouros, C., Chatzinotas S., Li, Y., Vucetic, B. and Ottersten, B. (2020a), "A tutorial on interference exploitation via symbol-level precoding: overview, state-of-the-art and future directions", *IEEE Communications Surveys and Tutorials*, Vol. 22 No. 2, pp. 796-839.

Li, B., Liu, Y., Zhang, A., Wang, W. and Wan, S. (2020b). "A survey on blocking technology of entity resolution", *Journal of Computer Science and Technology*, Vol. 35 No. 4, pp. 769-793

Li, B., Liang, R., Zhou, W., Yin, H., Gao, H. and Cai, K. (2021), "LBS meets blockchain: an efficient method with security preserving trust in SAGIN", *IEEE Internet of Things Journal*, available at: https://doi.org/10.1109/JIOT.2021.3064357.

Liu, Y., Lv, X. and Tang, Z. (2021), "The impact of mortality salience on quantified self behavior during the COVID-19 pandemic", *Personality and Individual Differences*, Vol. 180, p. 110972.

Liu, Y., Zhang, B., Feng, Y., Lv, X., Ji, D., Niu, Z., Yang Y., Zhao X. and Fan, Y. (2020), "Development of 340-GHz transceiver front end based on GaAs monolithic integration technology for THz active imaging array", *Applied Sciences*, Vol. 10 No. 21, pp. 7924.

Lv, X., Liu, Y., Xu, S. and Li, Q. (2021a), "Welcoming host, cozy house? The impact of service attitude on sensory experience", *International Journal of Hospitality Management*, Vol. 95, p. 102949.

Lv, Z., Qiao, L., Singh, A.K. and Wang, Q. (2021b), "Fine-Grained Visual Computing Based on Deep Learning", *ACM Transactions on Multimidia Computing Communications and Applications*, Vol. 17 No. 1s, pp. 1-19.

Lv, S. and Liu, Y. (2021), "PLVA: Privacy-Preserving and Lightweight V2I Authentication Protocol", *IEEE Transactions on Intelligent Transportation Systems*, doi: 10.1109/TITS.2021.3059638.

Ma, X., Zhang, K., Zhang, L., Yao, C., Yao, J., Wang, H., Jian, W. and Yan, Y. (2021), "Data-driven niching differential evolution with adaptive parameters control for history matching and uncertainty quantification", *SPE Journal*. doi: 10.2118/205014-PA.

McAleer, M. and Da Veiga, B. (2008), "Forecasting value-at-risk with a parsimonious portfolio spillover GARCH (PS-GARCH) model", *Journal of Forecasting*, Vol. 27 No. 1, pp. 1-19.

Ni, T., Yang, Z., Chang, H., Zhang, X., Lu, L., Yan, A., Huang, Z. and Wen, X. (2021), "A novel TDMA-based fault tolerance technique for the TSVs in 3D-ICs using honeycomb topology", *IEEE Transactions on Emerging Topics in Computing*, Vol. 9 No 2, pp. 724-734.

Niu, Z., Zhang, B., Wang, J., Liu, K., Chen, Z., Yang, K., Zhou, Z., Fan, Y., Zhang, Y., Ji, D., Feng, Y. and Liu, Y. (2020), "The research on 220GHz multicarrier high-speed communication system", *China Communications*, Vol. 17 No. 3, pp. 131-139.

Ouyang, L., Zhu, S., Ye, K., Park, C. and Wang, M. (2021), "Robust bayesian hierarchical modeling and inference using scale mixtures of normal distributions", IISE Transactions, pp. 1-13. doi: 10.1080/24725854.2021.1912440.

Pfeiffer, R.A.C. (2019), "Digital Economy, big data and competition law", *Market and Competition Law Review*, Vol. 3 No. 1, pp. 53-89.

Shi, J., Lu, Y. and Zhang, J. (2020), "Approximation attacks on strong PUFs", *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, Vol. 39 No. 10, pp. 2138-2151.

Sun, J. and Lv, X. (2021), "Feeling dark, seeing dark: Mind–body in dark tourism", *Annals of Tourism Research*, Vol. 86, p. 103087.

Sun, Y., Shi, Y. and Zhang, Z. (2019), "Finance big data: management, analysis, and applications", *International Journal of Electronic Commerce*, Vol. 23 No. 1, pp. 9-11.

Veselovsky, M.Y., Pogodina, T.V., Ilyukhina, R.V., Sigunova, T.A. and Kuzovleva, N. F. (2018), "Financial and economic mechanisms of promoting innovative activity in the context of the digital economy formation", *Entrepreneurship and Sustainability Issues*, Vol. 5 No. 3, pp. 672-681.

Wang, B., Zhang, B.F., Zou, F.C. and Xia, Y. (2021a), "A kind of improved quantum key distribution scheme", *Optik*, Vol. 235 No. 5, pp. 166628.

Wang, B., Zhang, B.F., Zou, F.C. and Xia, Y. (2021b), "A kind of improved quantum key distribution scheme", *Optik*, Vol. 235, p. 166628.

Wang, H., Han, Y., Fidrmuc, J. and Wei, D. (2021c), "Confucius institute, belt and road

initiative, and internationalization", *International Review of Economics & Finance*, Vol. 71, pp. 237-256.

Wu, C., Wu, P., Wang, J., Jiang, R., Chen, M. and Wang, X. (2020), "Critical review of data-driven decision-making in bridge operation and maintenance", *Structure and Infrastructure Engineering*, pp. 1-24. doi: 10.1080/15732479.2020.1833946.

Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A. and Wang, X. (2020), "Computer vision techniques in construction: a critical review", *Archives of Computational Methods in Engineering*, available at: https://doi.org/10.1007/s11831-020-09504-3.

Xue, X., Zhang, K., Tan, K.C., Feng, L., Wang, J., Chen, G., Zhao, X. and Yao, J. (2020a), "Affine transformation-enhanced multifactorial optimization for heterogeneous problems", *IEEE Transactions on Cybernetics*, available at: https://doi.org/10.1109/TCYB.2020.3036393.

Xue, X., Zhang, K., Tan, K.C., Feng, L., Wang, J., Chen, G., Zhao, X. and Yao, J. (2020b), "Affine transformation-enhanced multifactorial optimization for heterogeneous problems", *IEEE Transactions on Cybernetics*, doi: 10.1109/TCYB.2020.3036393.

Yang, Y., Liu, Y., Lv, X., Ai, J. and Li, Y. (2021), "Anthropomorphism and customers' willingness to use artificial intelligence service agents", *Journal of Hospitality Marketing & Management*, pp. 1-23. doi: 10.1080/19368623.2021.1926037.

Zhang, B., Ji, D., Fang, D., Liang, S., Fan, Y. and Chen, X. (2019a), "A novel 220-GHz GaN diode on-chip tripler with high driven power", *IEEE Electron Device Letters*, Vol. 40 No. 5, pp. 780-783.

Zhang, X., Wang, Y., Wang, C., Su, C., Li, Z. and Chen, X. (2019b), "Adaptive estimated inverse output-feedback quantized control for piezoelectric positioning stage", *IEEE Transactions on Cybernetics*, Vol. 49 No. 6, pp. 2106-2118.

Zhang, B., Niu, Z., Wang, J., Ji, D., Zhou, T., Liu, Y., Feng, Y., Hu, Y., Zhang, J. and Fan, Y. (2020a), "Four-hundred gigahertz broadband multi-branch waveguide coupler", IET Microwaves, *Antennas & Propagation*, Vol. 14 No. 11, pp. 1175-1179.

Zhang, Z., Liu, S. and Niu, B. (2020b), "Coordination mechanism of dual-channel closed-loop supply chains considering product quality and return", *Journal of Cleaner Production*, Vol. 248, p. 119273.

Zhang, J. and Shen, C. (2021), "Set-Based obfuscation for strong PUFs against machine learning attacks", *IEEE Transactions on Circuits and Systems. I: Regular Papers*, Vol. 68 No. 1, pp. 288-300.

Zhang, J., Shen, C., Su, H., Arafin, M.T. and Qu, G. (2021), "Voltage over-scaling-based lightweight authentication for IoT Security", *IEEE Transactions on Computers*, available at: https://doi.org/10.1109/TC.2021.3049543.

Zhou, Y., Tian, L., Zhu, C., Jin, X. and Sun, Y. (2020), "Video coding optimization for virtual reality 360-degree source", *IEEE Journal of Selected Topics in Signal Processing*, Vol. 14 No. 1, pp. 118-129.

Zikmund, G.W.(1994), *Business Research Methods*, 4th ed., The Dryden Press, New York, NY.

Zyskind, G. and Nathan, O. (2015),"Decentralising privacy: using blockchain to protect personaldata",2015 *IEEE Security and Privacy Workshops*, IEEE, San Jose, USA, pp. 180-184.