

The Impact of COVID-19 on Farmers' Economic Income in Hubei Province of China

Mingzhe E¹, Bin Zhao^{1*}, Jinming Cao²

¹School of Science, Hubei University of Technology, Wuhan, Hubei, China; ²School of Information and Mathematics, Yangtze University, Jingzhou, Hubei, China

ABSTRACT

This paper discusses the statistical measurement of the impact of COVID-19 major emergencies on farmers' economic income in Hubei Province. Hubei Province was selected as the object of analysis, and five data of total output value of agriculture, forestry, animal husbandry, fishery and per capita disposable income of farmers in Hubei Province from the first quarter of 2013 to the second quarter of 2020 were collected by using the Internet. Since all the collected data were macroeconomic data, these data were taken the logarithm to meet the economic significance.

The per capita disposable income of farmers was taken as the response variable, and the main factors affecting farmers' income were obtained by factor analysis.

Livestock husbandry and fishery industries were the main industries in Hubei Province. Then the score of factor analysis were taken as explained variable to establish regression model composed of influencing factors. This paper use the multiple linear regression, support vector regression to fitting and forecasting data, ARIMA model of time series analysis, introduced at the same time, through the AIC model choice, with the first quarter of 2013 to 2019 in the second quarter fitting training, backward prediction two quarters, and three or four quarter of 2019 compared with the real data, through to the predicted results of the sequence diagram and evaluation index model to compare the Mean Square Error (RMSE).

Three models predict per capita disposable income of farmers in the first and second quarter of 2020. It has been found that performance better ARIMA model in the model compare is worse than before, and three kinds of predicted values are higher than the real value of the model, showed the outbreak to the influence of the agricultural economy in Hubei province is serious. On this basis, taking into account the characteristics of geomorphic climate in Hubei province, the constructive suggestions are put forward.

Keywords: COVID-19; Multiple linear regression; Support vector regression; ARIMA model

INTRODUCTION

Thanks to the hard work of the whole nation, the epidemic prevention situation in China has been constantly improved, and the order of production and life has been quickly restored. In order to win the critical period of building a well-off society in all aspects and the decisive battle against poverty, the impact of COVID-19 epidemic on agriculture and rural economy should be scientifically studied and judged, and appropriate responses should be made. It is of great significance to ensure the well-being of a well-off society and the quality of poverty alleviation.

The impact of COVID-19 on agricultural economy mainly includes the impact on the supply of agricultural products, the impact on

livestock, poultry and aquaculture, the impact on the supply of agricultural materials and spring tillage preparation, and the impact on the international trade of agricultural products, etc. [1,2].

Research on the relationship between agricultural industrial structure and economic growth from the perspective of farmers' income level: Chen Kai [3] used the grey correlation method to analyse the relationship between rural household operating income and the output value of planting, forestry, animal husbandry and fishery in the Yangtze River Delta region. Yang Zhong-Na et al. [4] analysed the impact of agricultural structure and its changes on agricultural growth in southern Xinjiang from 2000 to 2011, and pointed out that agricultural industrial structure has restricted agricultural economic growth, indicating that the relationship

Correspondence to: Zhao B, School of Science, Hubei University of Technology, Wuhan, Hubei, China, Tel: +86 130 2851 7572; E-mail address: zhaobin835@nwsuaf.edu.cn

Received: January 27, 2021; **Accepted:** March 22, 2021; **Published:** March 29, 2021

Citation: Mingzhe E, Zhao B, Cao J (2021) The impact of COVID-19 on farmers' economic income in Hubei Province of China. J Med Diagn Meth. 10:322. 1.

Copyright: © 2021 Mingzhe E, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

between agricultural industrial structure and agricultural economic growth is spatially regional [5,6].

Therefore, it is very important to analyse the impact of the epidemic on agricultural economy, which can guide the direction of agricultural structural reform.

INDEX SELECTION

In this paper, the main income sources of farmers are selected as the analysis indicators, including agriculture, forestry, animal husbandry and fishery. Some data are shown in Figure 1.

<dbl>	<dbl>	<dbl>	<dbl>	<int>
217.80	8.83	391.30	85.8	2585
719.50	30.30	540.90	206.1	3991
2020.90	32.10	1065.50	363.5	6392
2505.15	122.00	1208.55	740.4	9682
228.50	7.19	393.20	89.0	2306
756.80	35.40	538.10	220.6	4513

Figure 1: Partial data diagram.

First, taking logarithm satisfies the economic significance, and then, through factor analysis, the internal relations between various variables are explored to mine the internal structure of the observed data and provide convenience for subsequent professional analysis.

ESTABLISHMENT AND SOLUTION OF THE MODEL

Factor analysis

Principle of factor analysis: Based on the results of correlation analysis, it is known that there is a certain correlation between the total output value of agriculture, forestry, animal husbandry and fishery and the per capita disposable income of rural residents. Factor analysis studies the relationship between dependent variables, so as to obtain a statistical analysis method of unobservable independent variables hidden among observable variables. Therefore, this paper selects the method of factor analysis to analyze the common factors affecting per capita disposable income, and constructs new variables according to the factor score, so as to eliminate the impact of the correlation of original data.

Factor analysis: Spearman correlation coefficient analysis of pre-data calculation in Hubei province was conducted in this paper, and the correlation coefficient matrix was visualized. The results are shown in Figure 2.

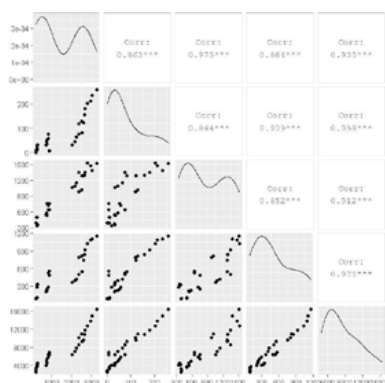


Figure 2: Spearman correlation coefficient matrix visualization of Hubei Province.

As can be seen from Figure 2, there is a strong correlation between all variables, with a correlation coefficient above 0.8 showing a strong correlation. Therefore, factor analysis is considered in this paper to eliminate the correlation between variables. KMO test and Bartlett spherical test were performed on the data, and the test results were shown in Table 1.

Table 1: KMO and Bartlett tests.

KMO	0.77
Bartlett test of sphericity	<0.01

As can be seen from Table 1, the KMO value of the data test is greater than 0.5, and the p value of Bartlett's spherical test is less than 0.05. Therefore, the test results can be considered as applicable to factor analysis.

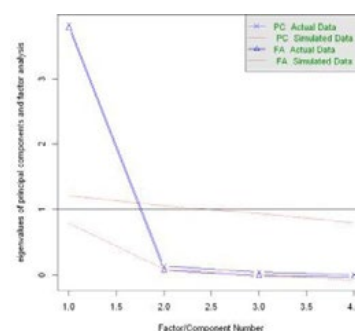


Figure 3: Factor analysis of parallel lithotripsy in Hubei Province.

The R programming language was used to draw a parallel gravel diagram (Figure 3) to determine the number of common factors. In Figure 3, we drew the gravel diagram for principal component analysis for comparison. If PCA principal component analysis was used in the diagram, the number of principal components to be selected from the PC curve was 1. According to the FA curve, the first two eigen values are above the corner and are greater than the mean value of the eigenvalues of the simulated data matrix for 100 times. Therefore, according to The Kaiser-Harris principle and the parallel analysis criteria, it can be considered that two common factors can be selected for factor analysis.

We determined that it was appropriate to select two common factors, and then carried out factor rotation on the data of Hubei Province, and drew oblique rotation factor result chart to determine the influencing factors represented by each factor. It can be concluded from the figure that factor 1 represents the driving factor of animal husbandry and factor 2 represents the driving factor of fishery. The time sequence diagram of model comparison and prediction is shown in Figure 4.

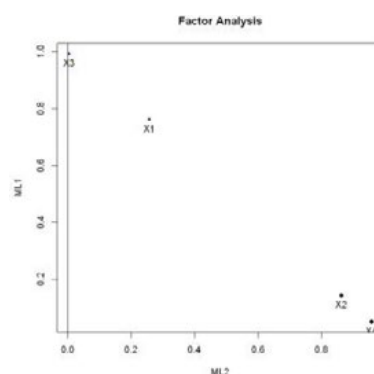


Figure 4: Two factor figure.

Multiple linear regressions

Regression methods can reflect the functional relationship between independent variables and dependent variables, while linear regression assumes that the functional relationship between independent variables and dependent variables is a linear functional relationship, dependent variables can be expressed as a linear combination of independent variables, and the regression of multiple independent variables becomes multiple regressions. The choice of linear regression is because the model is simple, intuitive and easy to calculate. The common method of model parameter estimation is the least squares regression estimation method to minimize the loss function. The linear regression model is expressed as:

$$y = X\beta + \varepsilon$$

Factor score: The results of the factor score table are shown in Table 2.

Table 2: Factor score table of Hubei Province.

	Factor1	Factor2
x_1	-0.005	0.517
x_2	0.193	0.015
x_3	-0.076	0.566
x_4	0.873	-0.112

The corresponding expressions of the two factors can be obtained by converting the data in Table 2 into the factor score function:

$$f_1 = -0.005x_1 + 0.193x_2 - 0.076x_3 + 0.873x_4$$

$$f_2 = 0.517x_1 + 0.015x_2 + 0.566x_3 - 0.112x_4$$

According to the information in Table 2, the common factor 1 represents the animal husbandry driving factor; Common factor 2 represents the fishery driving factor. According to the above analysis, it can be known that for Hubei Province, the main influencing factors of agricultural structure adjustment on farmers' income include driving the overall economic development led by animal husbandry and vigorously developing animal husbandry, among which the first influencing factor is the most important.

Linear regression: After finding common factors through factor analysis in Section 3.1 and dimensionality reduction of the original data set, the multiple linear regression model of per capita disposable income of farmers and two common factors was established by using multiple linear regression method.

1. Multiple regression models: In this paper, R language programming was used to carry out multiple regressions modeling of farmers' income and its two common factors in Hubei Province, and the regression equation was obtained as follows:

$$y = 0.36f_1 + 0.29f_2 + 4.90$$

2. Model test:

- **R square test:** The R square test can judge the degree of interpretation of independent variables to dependent variables. The closer R square is to 1, the greater the degree of

interpretation of independent variables to dependent variables will be. It is known from the table 3 that the common factors of the multiple regression models have a good interpretation of farmers' income.

Table 3: R squared test value.

Province	R squared
Hubei	0.9584

- **DW test:** The DW test is used to test the independence between the observed values. The closer the VALUE of the DW test is to 2, the more independent the observed values are. As shown by the table 4, the value of the common factor of the multiple regression models is close to 2, so it can be considered that the common factor has good independence.

Table 4: RDW test value.

Province	RDW
Hubei	1.4695

- **F test:** F test is a test method to determine whether there is a linear relationship between independent variables and dependent variables. When the p value of F test is less than 0.05, it is considered that there is a linear relationship between agriculture, forestry, animal husbandry, fishing and farmers' income. It can be seen from the table 5 that the p value of F test is far less than 0.05, and it can be considered that there is a linear correlation between the common factor of the multiple linear model and farmers' income.

Table 5: p value of F test.

Province	p value
Hubei	<0.01

- **Multicollinearity test:** If there is a correlation between independent variables, the model is considered to have multiple common linearity. In this paper, VIF index is selected to test whether there is multicollinearity among independent variables. When VIF value is less than 10, it can be considered that there is no multicollinearity among independent variables in this multiple regression model. It is known from the table 6 that the VIF test values of the linear model are all less than 10, so it can be considered that there is no multi-collinearity between the common factors selected in Section 3.1.

Table 6: VIF test value.

Province	VIF
Hubei	3.48

- **Model prediction:** In order to study the impact of the epidemic on agricultural economic income, we used the historical data from 2013 to 2019 to establish a multiple linear regression model, and used this model to predict the economic income of farmers in the first and second quarters of 2020, and compared

and analyzed the predicted data with the actual data, so as to analyze the impact of the epidemic on farmers. We use the mean square error.

Table 7: Linear regression analysis for the first two quarters of 2020.

Province	Forecast for the first quarter of 2020	Real value for the 1 st quarter of 2020	Forecast for the 2 nd quarter of 2020	Real value for the 2 nd quarter of 2020	RMSE
Hubei	8.32	8.31	8.86	8.76	0.072

It can be seen from the table 7 that the predicted value of Hubei Province in the first and second quarters of 2020 is higher than the true value based on the time series prediction of the data from 2013 to 2019 without the influence of epidemic as the observed value, indicating that the epidemic has a relatively large impact on the disposable income of farmers in Hubei Province, and the epidemic will reduce the economic income of farmers.

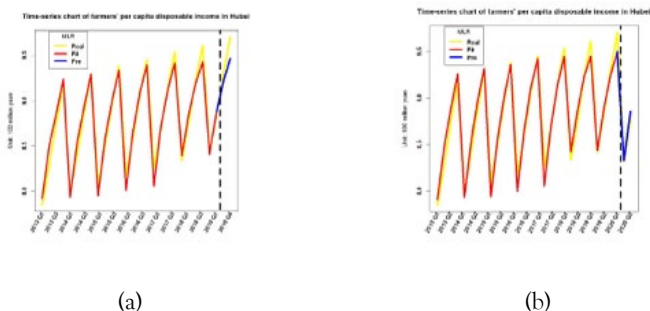


Figure 5: Comparison and prediction sequence diagram of multiple linear regression models; (a) Comparison time series; (b) Forecast time series.

It can be seen from Figure 5 that the predicted value in the third and fourth quarters of 2019 in Hubei Province is lower than the real value, indicating that the farmers' income is rising steadily, while the predicted value in the first and second quarters of 2020 is higher than the real value, indicating that the rural economy in these two quarters is seriously affected by the epidemic.

Support vector regression

Model building: In a Support Vector Regression (SVR) model, a sample set is given $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, then getting a regression model $f(x)$ and y as similar as possible, m and b is the model $f(x) = \omega^T x + b$, make $f(x)$ and ω as similar, ω and b is the model parameters.

For samples (x, y) , traditional regression models usually calculate the loss based on the difference between model output $f(x)$ and real value $f(x)$. The losses are zero if and only if $f(x)$ and y are exactly equal. On the contrary, The SVR model assumes that we can tolerate an error of ϵ at most between $f(x)$ and y , and the loss is calculated only when the absolute value of the difference between $f(x)$ and y is greater than ϵ [7].

As shown in Figure 6 below, the points inside the blue strip have no loss, while the points outside have loss, and the loss is the length of the red line.

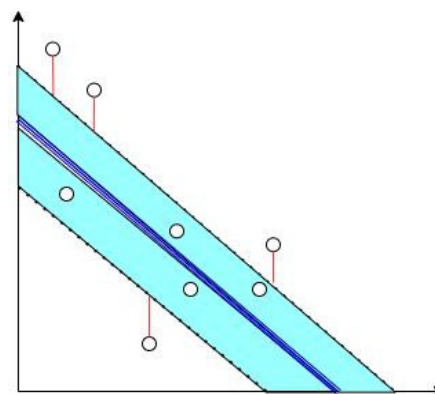


Figure 6: Support vector regression diagram.

Solution and test: First used in the first quarter of 2013 to 2019 in the second quarter fitting training, backward prediction two quarters, and three or four quarter of 2019 compared with the real data, compared with other models, in the first quarter of 2013 to 2019 in the fourth quarter of data forecast in 2020 first and second quarters, to examine the rationality of the model, analyze each model of residual error as shown in figure 6, six models are reasonable.

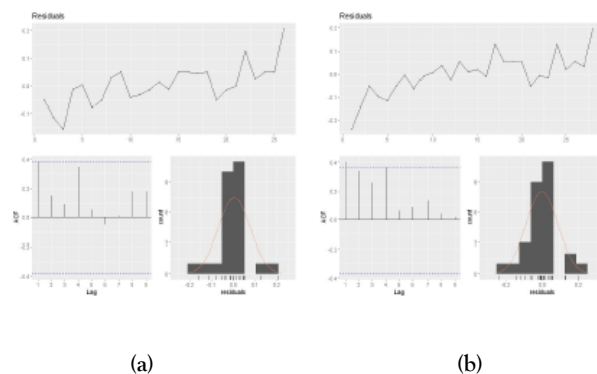


Figure 7: Model comparison with forecast residual graph; (a) Model comparison residuals; (b) The model predicts residuals.

Analysis and interpretation of model results: Time sequence diagram of model comparison and prediction is shown in Figure 7.

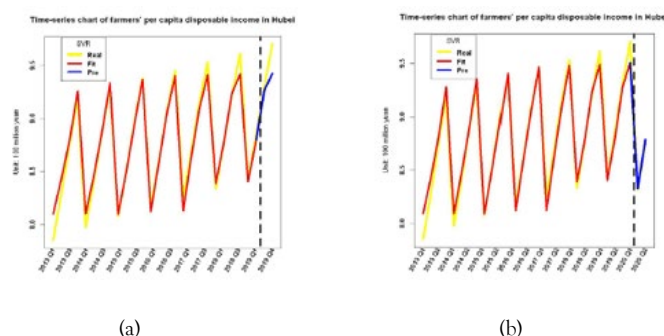


Figure 8: Comparison and prediction sequence diagram of SVR models; (a) Comparison time series; (b) Forecast time series.

It can be seen from the results that the predicted values in Hubei Province are all higher than the real values, indicating that the income of farmers in Hubei Province is seriously affected by the epidemic (Figure 8).

ARIMA model

Model Building: This module uses the autoregressive moving

average model (ARIMA model), which is the most commonly used model in time series prediction. We also call it the differential autoregressive moving average model. In this model, AR, P, MA and Q represent autoregressive term, autoregressive term, moving average term and moving average term respectively, and D represents the number of difference we make when the time series is stable.

Where respectively represent the autoregressive coefficient polynomial and the moving smoothing coefficient polynomial of the stationary reversible time series model.

Before establishing the ARIMA model, we first determine the order of the model, i.e. the values of P and Q, and select the appropriate order to achieve the relative optimization of the final fitting model. In the time series model, it is necessary to order the model, and we generally use AIC criterion and BIC criterion.

Where AIC is the weighting function:

$$AIC = -2\ln(L) + 2K$$

In order to achieve the optimal model, we should try to choose the minimum AIC value. In this paper, we use the data from the first and second quarters of 2013 to 2019 to build the ARIMA model, and then predict the per capita disposable income in the first and second quarters of 2020 and compare it with the real value.

Solution and test: After the establishment of the model, it is necessary to evaluate the established time series model. First, we need to test the fitting effect of the time series model through the stationary R-square. Then the white noise test was carried out for the residuals. We can use Q test to determine whether the residual is white noise.

Based on time series analysis, in this paper, the farmers of disposable income, the integrated test, first of all to log processing of raw data, then a stabilized and zero mean treatment of the ARIMA model was established, and the one quarter of Hubei province in 2020 to predict the disposable income of farmers, analyzes the influence of the outbreak.

The null hypothesis of the Q test is: $H_0 : \rho_1 = \rho_2 = \dots = \rho_n = 0$ and the alternative hypothesis is: H_1 : one of $\rho_i (i=1, 2, \dots, n)$ is non-zero at least. Under the condition that the null hypothesis H_0 is established, the statistics $Q = T(T+2) \sum_{k=1}^n \frac{r_k^2}{T-k} \sim \chi^2_{s-n}$ are established. Where T is the number of samples n denotes the number of position parameters in the model. When the p value obtained by Q test is less than 0.05, the null hypothesis is rejected. It is believed that the model does not fully recognize the data. If a p value is greater than 0.05, the model is accepted as recognizing the data well.

In addition, we made ACF and PACF graphs of residuals to judge whether the residuals are white noise. If ACF and PACF of each order are less than the critical value of the test, the residual is considered to be white noise, that is, the model identifies the data well. Otherwise, it is considered that the model is not fully identified. The result is shown in Figure 9:

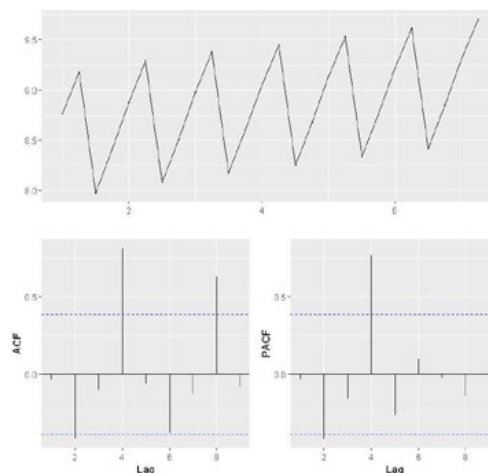


Figure 9: Peasant per capita disposable income time series model residuals ACF and PACF.

By analyzing the ACF and PACF graphs of the time-series model residuals corresponding to all variables in Hubei Province, it can be found that they are all less than the critical value of the test. Therefore, the residuals can be considered as white noise. That is to say, the model obtained in this paper can identify our data well. So we can accept these temporal models.

Analysis and interpretation of model results: Through the obtained time series model, we can predict the per capita disposable income. Figure 10 below shows the time series diagram of the per capita disposable income of farmers in Hubei Province.

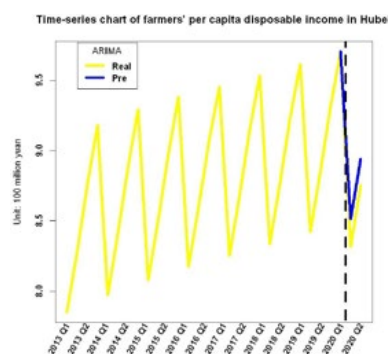


Figure 10: Timing chart of farmers' per capita disposable income.

It can be seen from Figure 10 that the predicted value of Hubei Province in the first and second quarters of 2020 is higher than the true value based on the time series prediction based on the data from 2013 to 2019 without the impact of the epidemic, indicating that the epidemic has a relatively large impact on the disposable income of farmers in Hubei Province.

Comparative analysis of models

The mean square errors of comparison and prediction of all models are shown in Table 8 below.

Table 8: RMSE comparison table for each model.

	Province	ARIME	LM	SVR
Comparison	Hubei	0.002	0.179	0.202
Prediction	Hubei	0.192	0.072	0.018

As can be seen from Table 8, in comparison and prediction of Hubei Province, the performance of ARIMA model which performed better before is worse than before, and the predicted values of the

four models are all higher than the real values, indicating that the income of farmers in Hubei Province is seriously affected by the epidemic.

CONCLUSION AND RECOMMENDATION

Based on the modelling and analysis in this paper, we successively put forward suggestions on the impact of agricultural structure adjustment on farmers' income in Hubei Province.

Strengthen macro-control

Although we strongly advocate market agriculture and emphasize the dominant position of farmers in the adjustment of agricultural industrial structure, the government should strengthen macro-control in the face of poor income increase of farmers and excessive pursuit of immediate interests, which impair industrial adjustment. In the process of adjusting agricultural industrial structure in Hubei Province, all the work should be carried out with increasing farmers' income as the goal, to support farmers' business activities in accordance with the industrial structure adjustment as much as possible, and actively play the function of macro-control.

Develop aquaculture vigorously

The low proportion of aquaculture in the primary industry is an important manifestation of the irrational agricultural industrial structure. Hubei province has rich water resources and unique advantages for aquaculture, but the proportion of fishery in the agricultural industry in Hubei province is only higher than forestry. Therefore, we should pay more attention to aquaculture and vigorously develop aquaculture with local characteristics, such as crayfish in Qianjiang and crab in Xianning, so as to make it the key point to promote farmers' income.

Strengthen scientific training

Local governments and agricultural departments should stick to the central idea of "developing agriculture through science and technology", persist in improving farmers' quality through "science and technology training" as a media platform, and help farmers solve various technical problems they may encounter in production, so as to improve farmers' agricultural planting techniques. Through the combination of "centralized teaching, on-site teaching and individual guidance", the innovative mode of agricultural cultivation is promoted, and the prevention and control analysis of possible insect disease disasters is carried out, so that farmers can quickly master skills. The government's science and technology

training platform plays an important role in strengthening farmers and radiating the surrounding economic development.

Agricultural economy

Combined with China's special rural conditions, coupled with the lack of implementation of relevant national benefits to farmers, their own land, lack of funds, single information channels, resulting in farmers in the process of agricultural industrial structure adjustment in the face of the market alone in the case of higher costs, limit the increase of income level. In the face of this situation, it is very important to gradually develop a new form of cooperative organization to help the separated and independent farmers establish connections with the large market.

REFERENCES

1. Dong Li. Influence of Covid-19 epidemic on agricultural and rural economy and countermeasures. *Contem Rural Fin Econo*. 2020;5:5-6.
2. Hong-Yu Zhang, Ling-Xiao Hu. The impact of the Covid-19 epidemic on agricultural and rural economies. *Chinese rural finance*. 2020;8:31-34.
3. Kai C. Adjustment of agricultural structure, agricultural versatility and changes of farmers' income: An empirical study of Jiangsu, Zhejiang and Shanghai regions in the Yangtze River Delta. *Econo Iss*. 2011;11:82-86.
4. Zhong-Na Y, Ji-Jun T. The influence of agricultural structure change on agricultural economic growth in southern Xinjiang. *Agricultural resources and regionalization in China*. 2014;35(3):85-92.
5. Yu-Lu Chen. The impact on China's economy has been limited. *Fulcrum*. 2020;3:12.
6. De-Hua W. Discussion on the economic impact of COVID-19 epidemic and the ideas of fiscal hedging policy. *China's finance*. 2020;10:22-24.
7. Zhi-Hua Z. *Machine Learning*. Peking: Tsinghua University Press. 2016.01.
8. Tan FF, Lu Z. Study on the interaction and relation of society, economy and environment based on PCA-VAR model: As a case study of the Bohai Rim region, China. *Ecol Indic*. 2015;48:31-40.
9. Chen RD, Cao D, Yu Q. Nonlinear VaR model of FX options portfolio under multivariate mixture of normal distributions. *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice*. 2009; 29(12):65-72.
10. Fan K, Nie G. Study on china's import and export growth rate based on var model. *Lecture notes in electrical engineering*, 2013;2:295-304.