THE IMPACT OF COVID-19 ON FARMERS' ECONOMIC INCOME IN HUBEI PROVINCE OF CHINA DURING THE COVID-19 EPIDEMIC

Mingzhe E[1], Bin Zhao[1]*, Jinming Cao[2]

[1]School of Science, Hubei University of Technology, Wuhan, Hubei, China.

[2]School of Information and Mathematics, Yangtze University, Jingzhou, Hubei, China.

E-mail: zhaobin835@nwsuaf.edu.cn

ABSTRACT

This paper discusses the statistical measurement of the impact of COVID-19 majoremergencies 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.

1. 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 analyze 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] analyzed 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 between agricultural industrial structure and agricultural economic growth is spatially regional [5][6].

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

2. 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.
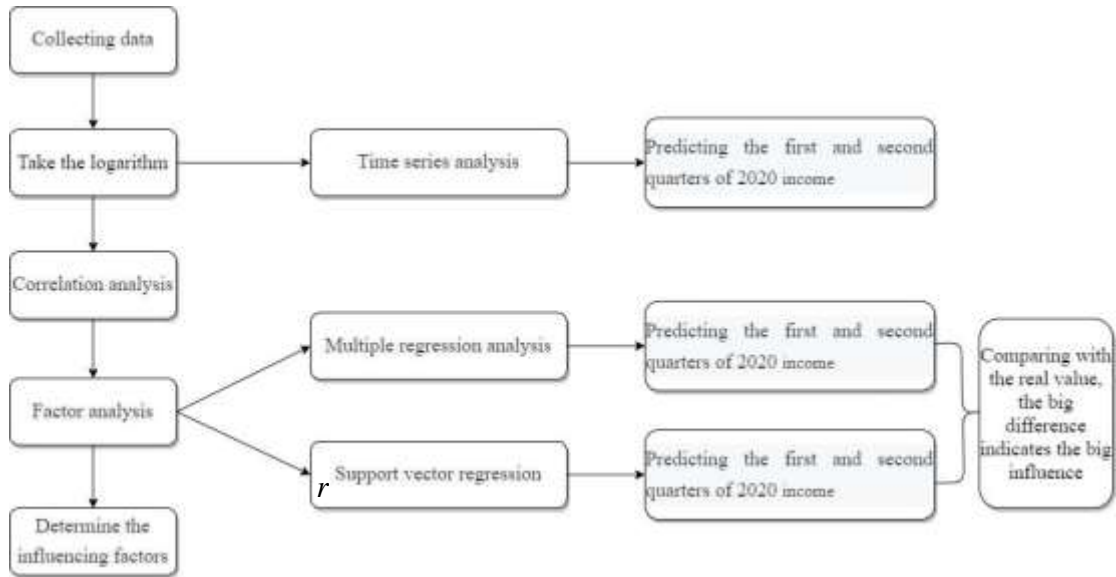
Figure 1: Partial data diagram

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.

3. Establishment and solution of the model The flow chart of the overall thinking is as follows:

<dbl>	<dbl>	<dbl>	<dbl>	<int>
217.80	6.83	391.30	85.6	2565
719.50	30.30	540.90	208.1	3991
2020.90	32.10	1065.50	363.5	6392
2585.15	122.00	1296.55	748.4	9692
228.50	7.18	393.20	88.0	2908
756.80	35.40	538.10	220.8	4513

3.1 Factor analysis

3.1.1 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. For the 5 variables of total output value of agriculture, forestry, animal husbandry and fishery and per capita disposable income of rural residents, we used principal component method to conduct factor analysis:

（1）Calculate the correlation coefficient matrix

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$ Where r_{ij} is the correlation coefficient, $r_{ij} = r_{ji}$

（2）Feasibility of factor analysis of data: KMO test and Barrett ball test were performed on the data when the KMO value was greater than 0.5. When P-value of Bartlett test is less than 0.05, it is considered that factor analysis can be carried out.

（3） Carry out EFA analysis to determine the number of common factors, solve the characteristic equation $|I - R| = 0$ and solve the eigenvalues, and arrange them in order from large to small to draw parallel gravel diagram. The current n eigenvalues are above the corner, and they are all greater than the mean value of the eigenvalues of the simulation data matrix for 100 times, and the number of common factors can be determined as n.

（4）Factor rotation: The oblique rotation is adopted to extract the factor, which is beneficial to better explain the practical significance of the factor.
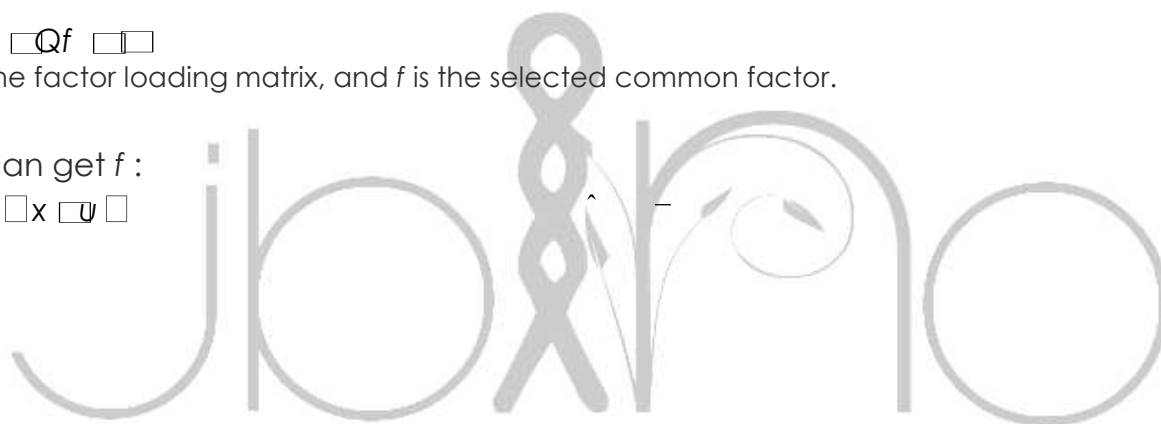
（5）According to the common factor extracted by oblique rotation, the factor score model is used to calculate the factor score. The factor score model represents variables as linear combinations of factors with error terms:

$$x = \mu + Qf + \epsilon$$

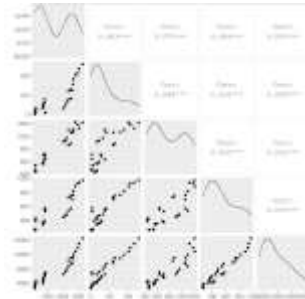Q is the factor loading matrix, and f is the selected common factor.

we can get f :

$$\hat{f} = B(x - \mu)$$

Where $\hat{\beta}$ is the estimated coefficient calculated by various regression methods.

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.

3.1.2 Factor analysis

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 (FIG. 3) to determine the number of common factors. In FIG. 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 eigenvalues 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.
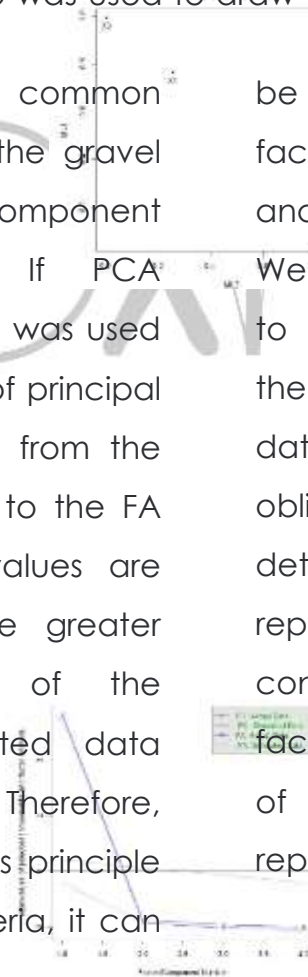
Figure 3:   Two factor figure

3.2 Multiple linear regression

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 regression. 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:

$y \ \square X \square \square \square \square$

3.2.1 Factor score

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

Table 2: Factor score table of Hubei Province

	Factor1 x1	-0.005 x2
0.193 x3		-0.076
x4		0.873
	Factor2 0.517 0.015 0.566	

-0.112The corresponding expressions of the two factors can be obtained by converting the

data in Table 2 into the factor score function:

$$f_1 = -0.005x_1 + 0.193x_2 - 0.076x_3$$
$$+ 0.873x_4 \quad f_2 = 0.517x_1 + 0.015x_2$$
$$+ 0.566x_3 - 0.112x_4$$

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.D

Table 5: *R* squared test value

Province	*R* squared
Hubei	0.9584

（2）DW test

3.2.2 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 model

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

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

2. Model test

（1）*R* square test

The *R* square test can judge the degree of interpretation of independent variables todependent 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 that the common factors of the multiple regression model have a good interpretation of farmers' income.

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, the value of the common factor of the multiple regression model is close to 2, so it can be considered that the common factor has good independence.

Table 6: RDW test value

Province	DW
Hubei	1.4695 （3） *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 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 7: P value of F test

Province	P value
Hubei	<0.01 （4） 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 that the VIF test values

of the linear model are all less than 10, so it can be considered that there is no multicollinearity between the common factors selected in Section 3.1.

Table 8: VIF test value

| Province	VIF Hubei
	3.48

3. 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 first quarter of 2020	Forecast for the 2th quarter of 2020	Real value for the 2th quarter of 2020	RMSE
				8.76	0.072
Hubei	8.32	8.31	8.86		

It can be seen from the table 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.

The time sequence diagram of model comparison and prediction is shown in Figure 5:
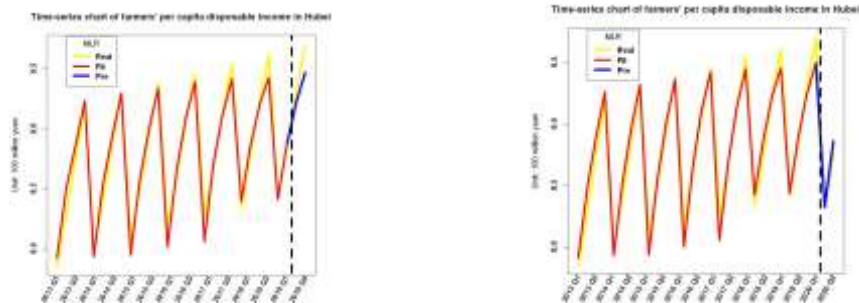
（a）Comparison time series   （b）Forecast time series **Figure 5:**
Comparison and prediction sequence diagram of multiple linear regression models

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.

3.3 Support vector regression

3.3.1 Model Building

In a support vector regression (SVR) model, a sample set is given $D = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\}$, then getting a regression model $f(x) = \omega^T x + b$, make $f(x)$ and y as similar as possible, 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 y. The loss is 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.
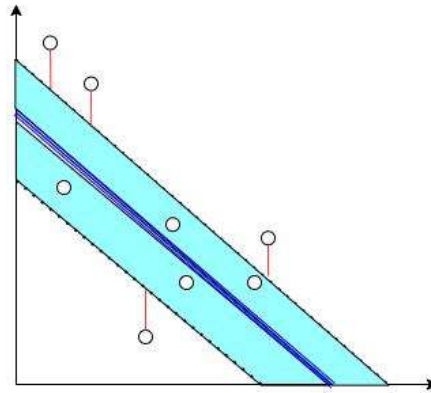
Figure 6: Support vector regression diagram The loss function of SVR model is measured as:

$$err(x_i, y_i) = \begin{cases} ,|y_0 - \omega \cdot \phi(x_i) - b| \le \varepsilon \\ |y_i - \omega \cdot \phi(x_i) - b| - \varepsilon ,|y - \omega \cdot \phi(x) - b| > \varepsilon \end{cases}$$

The objective function of SVR model is:

$$\min \frac{1}{2}\|\omega\|^2 \quad s.t. |y_i - \omega \cdot \phi(x_i) - b| \le \varepsilon (i = 1,2,...,m)$$

Thus, the SVR problem is written as:

$$\min_{\omega,b} \frac{1}{2}\|\omega\|^2 + C \sum_{i=1}^{m} l_\varepsilon (f(x_i) - y_i)$$

C is the regularization constant, l_ε is:

$$l_\varepsilon(z) = \begin{cases} 0, if\ |z| \le \varepsilon \\ |z| - \varepsilon, otherwise \end{cases}$$

Introducing relaxation variables ξ_i and $\hat{\xi}_i$, rewrite the above formula:

$$\min_{w,b,\xi_i,\hat{\xi}_i} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \hat{\xi}_i) \quad s.t.\ f(x_i) - y_i \le \varepsilon + \xi_i, \xi_i \ge 0, \hat{\xi}_i \ge 0, i = 1, 2, ..., m$$

Let's introduce the lagrange multipliers , The following Lagrangian function is obtained:

$$L(\omega,b,\alpha_i,\alpha_2,\xi,\xi,\mu,\mu) = \frac{1}{2}\|\omega\|^2 + C\sum(\xi + \xi) +$$

$$\sum_i^m (y_i - \omega(x_i) - b) - \sum_i^m (y_i - \omega(x_i) - b) - \sum_i^m \mu_i \xi_i - \sum \mu$$

Where _____ are Lagrange coefficients. Finally, using SMO algorithm to find the corresponding ____ , then we can get the coefficient of our regression model __ .

3.3.2 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 7, six models are reasonable.
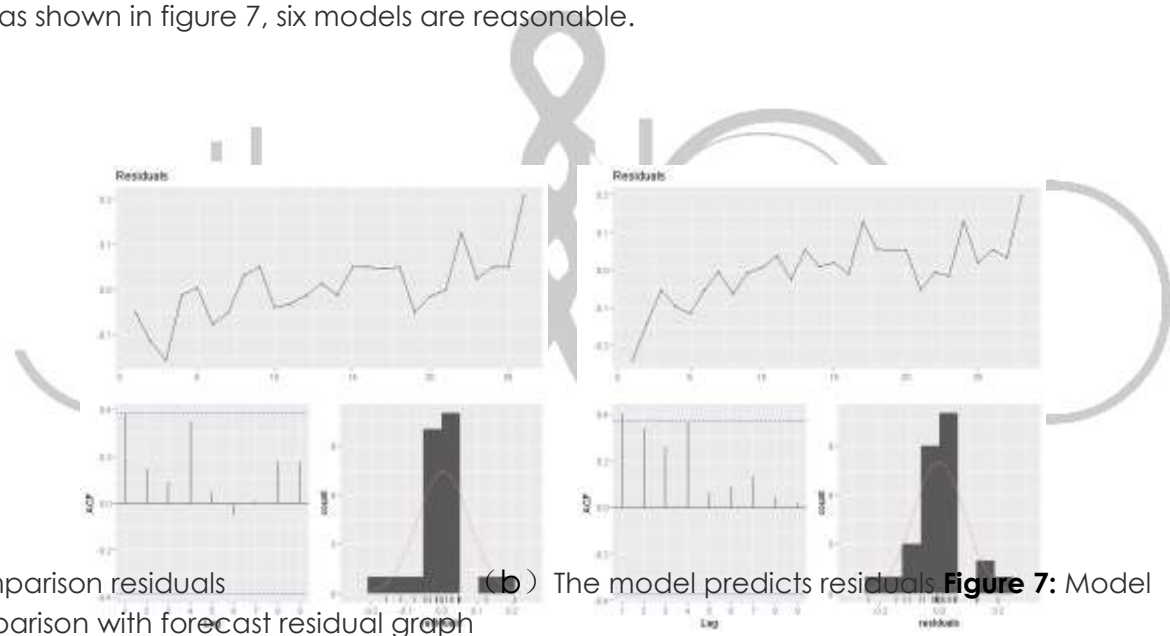
Model comparison residuals   （b）The model predicts residuals **Figure 7:** Model comparison with forecast residual graph

3.3.3 Analysis and interpretation of model results
Time sequence diagram of model comparison and prediction is shown in Figure 8.
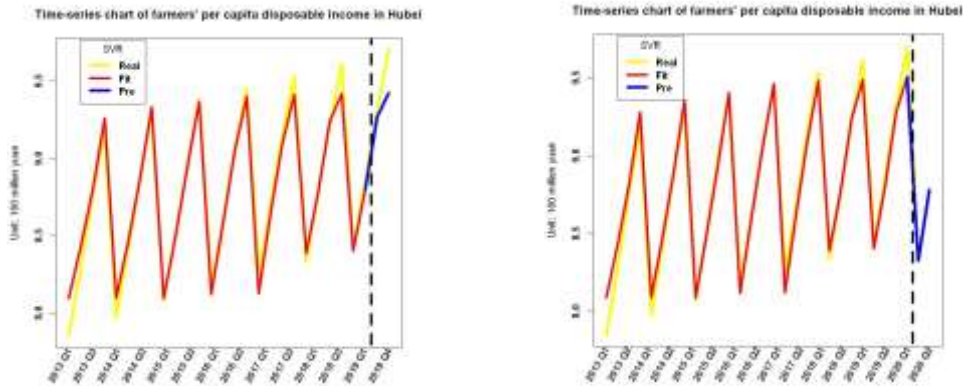
（a）Comparison time series   （b）Forecast time series **Figure 8:** Comparison and prediction sequence diagram of SVR models

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.

3.4 ARIMA model

3.4.1 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. The structure of ARIMA(p,d,q) model is shown as follows:

$$\Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t$$
$$E(\varepsilon_t) = 0, var(\varepsilon_t) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \quad E(\varepsilon_t) = 0, \forall s \neq t$$

Where   th   t   h $_h$ and   （) $_h$

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 ⊏⊐2In(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.

3.4.2 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 $\rho_1 = \rho_2 = \cdots = \rho_s$ and the alternative hypothesis is : : one of $\rho_i\ (i=1,2\ \text{is non-zero at least})$. Under the condition that the null hypothesis is established, the statistics $Q = T(T+2)\sum_{i-1}\dfrac{r_i^2}{T-k}$ 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:
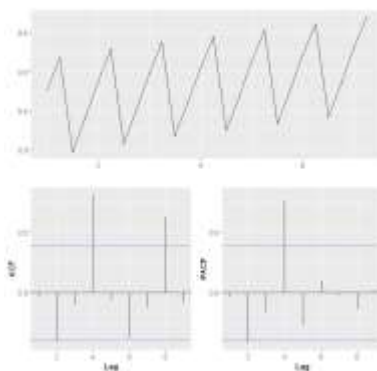

Figure 9: Peasant per capita disposable income time series model residuals ACF and PACF

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.

3.4.3 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.
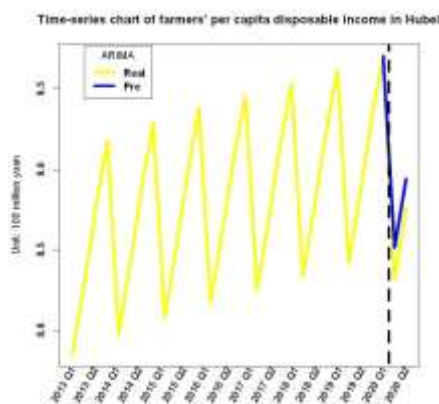

Figure 10: Timing chart of farmers' per capita disposable income

It can be seen from Fig. 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.

3.5 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.

4. Conclusions and Recommendations

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

（1）**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. （2）**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 offishery 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.

（3）**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. （**4**）**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.

Bibliography：

[1]   Dong Li, Influence of Covid-19 epidemic on agricultural and rural econom y and countermeasures[J]. Contemporary Rural Finance and Economics, 202 0(05)：5-6.

[2]   Hong-Yu Zhang, Ling-Xiao Hu. The impact of the Covid-19 epidemic on agricultural and rural economies[J]. Chinese rural finance, 2020(08)：31-34.

[3]   Kai Chen. 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[J]. Economic issue, 2011 (11)：82-86.

[4]   Zhong-Na Yang, Ji-Jun Tang. The influence of agricultural structure change on agricultural economic growth in southern Xinjiang [J]. Agricultural resources and regionalization in China, 2014, 35(03)：85-92.

[5]   Yu-Lu Chen. The impact on China's economy has been limited [J]. Fulcrum, 2020(03)：12.

[6]   De-Hua Wang. Discussion on the economic impact of COVID-19 epidemic and the ideas of fiscal hedging policy[J]. China's finance, 2020(10)：22-24.

[7]   Zhi-Hua Zhou. Machine Learning[M]. Peking: Tsinghua University Press. 2016.01.