Research Article

Empirical Study on the Grain Output Based on Regression Analysis

Jiahao Xu, Sai Tang, Pengyan Li, and Hexu Zhang

Harbin Institute of Technology, China

Correspondence should be addressed to Sai Tang; tangsai86@126.com

Received 27 June 2022; Revised 25 July 2022; Accepted 4 August 2022; Published 28 September 2022

Academic Editor: Yuan Li

Copyright © 2022 Jiahao Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on a literature review of influencing factors and forecasting methods for grain production, the empirical analysis of the influencing factors of China’s grain output is performed using the full subset regression method, the ridge regression method, and the LASSO regression method. The results show that (1) the increase in the sown area of grain crops is the main reason for the increase in grain output, (2) the use of agricultural fertilizers and the increase in rural electricity consumption are the driving factors for the increase in grain output, (3) the impact of total power of agricultural machinery is limited, and (4) natural disasters have a certain negative impact on food production.

1. Introduction

Food is not only the basis for the people’s survival but also the basis for the development of a country. At present, China’s economic activities are stable, but the external environment is complex and severe, the economy is facing downward pressure, and risks and difficulties have increased significantly. If there is a problem with the supply of food and important agricultural products, it will not only lead to rising prices, which will cause economic development to fall into a passive situation of downward growth and rising prices, but will also affect social stability. Therefore, it is particularly important to deepen agricultural supply-side reforms and increase grain production capacity. There are some of the key strategies for lowering farm produce production costs, and raising farmer income is to improve agricultural productivity. The level of farmers’ income is influenced by their motivation and the safety of the country’s food supply. The primary objective of the structural reform of the agricultural supply side, as stated in 2017’s No. 1 Central Document, is to raise farmers’ incomes and assure an adequate supply. The topic of peasant revenue and reducing the income disparity among urban and rural inhabitants as a “14th Five-Year” period in 2019 and 2021 was discussed in the No. 1 Central Document once more [1]. Generally, everyone understands that land is where farmers get the majority of their revenue. When grain prices are constant, increasing cost effectiveness and lowering production expenses are practical and efficient strategies to increase farmers’ revenue [1, 2].

Since the reform and opening up, especially since the implementation of the household coproduction contract responsibility system, due to the continuous optimization of agricultural-related policies, the continuous improvement of agricultural technology, the deepening of market-oriented reforms, and the dividends released by China’s overall macro-economic development, China’s food production has been greatly improved [3]. Taking 2019 as an example, the country’s total grain output was 66.384 million tons, achieving “sixteen consecutive harvests.” However, due to the influence of uncertain factors such as changes in the industrial structure, increasing restrictions on grain production factors, and fluctuations in the international grain market, the space for further increase in China’s grain output has gradually narrowed. Therefore, in order to “ensure basic self-sufficiency of grain and absolute security of rations,” it is necessary to clarify important factors affecting grain production, determine effective ways to further increase grain production, continue to do a good job of stabilizing agricultural production and ensuring supply and increasing farmers’ income, promote high-quality
agricultural development, and maintain harmony and stability in rural society [4, 5].

Based on the summary of previous research on influencing factors and prediction methods of grain production, this paper uses the full subset regression method, the ridge regression method, and the LASSO regression method to conduct empirical analysis of the main influencing factors of China’s grain output from 1991 to 2018.

2. Literature Review

The research on the influencing factors of grain output is mainly carried out from six aspects. First, agricultural resources such as arable land and water are hard constraints on food security. Xiaoshi et al. analyzed the relationship between changes in the quantity and quality of cultivated land and food production and argued that with the economic development and industrialization process, a large amount of agricultural land was deagriculturalized, which had negatively affected food security [1, 6, 7]. Ahmed and Melesse and Han et al. believed that cultivated land resources were the most basic material conditions for agricultural production, and changes in their quantities directly affect food production and food security [7, 8]. Second, population size and population structure will affect food security. Wei et al. and Feng believed that with population growth and consumption expansion, China’s future arable land size and per capita arable land area would further decline, and the per capita food consumption level and total food demand would further increase [2, 9]. Yuqiu and Lei believed that the adjustment of China’s future population structure would further reduce the growth rate of total food demand [4]. To some extent, changes in population structure have a positive impact on food security. Xiaoshi et al. and Taotao et al. explored the impact of demographic changes on China’s food security from the perspective of supply and demand. The results showed that with the advancement of population aging and population urbanization, Xiaoshi et al. were of the opinion that China’s food security had suffered a lasting negative impact [1, 10]. Compared with unhealthy workers, healthy workers have advantages in “rational” choice. On the one hand, as an input factor, healthy human capital participates in labor and forms a reasonable total allocation with other production factors [7]. On the other hand, due to the strong adaptability of labor intensity, it can avoid the unconditional excessive. A study by Xiaoshi et al. and Wei et al. pointed out that the impact of human capital on agricultural production efficiency would change with the scale of cultivated land. However, there are also some literatures that the contribution of human capital in agricultural production is not significant [1, 2]. Jingbo and Yuan and Xiang and Zhong found that the human capital had a positive role in promoting rural economic growth, but the contribution rate was low [3, 11]. Even a few scholars like Sanusi and Singh and Xuejiao and Haifeng have shown that the effect of human capital on rural economic growth was not significant and even sometimes played a negative role. Third, the relationship between urbanization and food security is studied [12, 13]. Wang conducted an empirical study on the relationship between urbanization and food security using panel data from 31 provinces and cities from 1997 to 2015. The results showed that the development of urbanization had an adverse effect on food security [14]. Cai took Henan Province as the research object and found that rapid urbanization had imposed constraints on food production, and a large number of rural laborers had moved to cities. The expansion of cities and towns had caused large areas of farmland to be converted to nonagricultural land, which had led to increased food security pressure [15]. Fourth is how agricultural technological progress affects food security. Xiaoshi et al. and Du believed that agricultural technological progress had a great impact on food security, and it was necessary to increase investment in agricultural technological progress and increased the use of advanced agricultural technologies to improve the level of food security [1, 16]. Fifth, impact of climate change on food security has been considered. Chen and Xie and Chih-Ming believed that climate disasters would severely affect the balance of food production and food production systems and that an agricultural meteorological disaster defense system needed to be built to improve China’s food security level [17, 18]. Chih-Ming and Qiu et al. believed that climate warming would threaten China’s food security. It would also change the global food trade pattern and increase factors of instability [18, 19]. In addition to the above factors, the existing literature has also studied from more micro perspectives such as how drought, the use of bio-mass energy, and genetic modification affect food security.

From the perspective of the main influencing factors, the research mainly focuses on regression and grey correlation analysis. For example, Yin et al. used the data from 1991 to 2005 and 2006 to 2010 to conduct grey prediction and grey correlation analysis. The results showed that the irrigated area, agricultural production materials, agricultural product prices, and grain sown were the main factors which affected grain production [20]; Zai et al. used stepwise regression to perform a regression analysis on 12 factors affecting grain production in Henan Province and determined that the main influencing factors were fertilizer application, pesticide application, and agriculture were the three expenses for fertilizer application, pesticide application and agriculture technology and the planting area [21]; Qiling and Zhang also used gradual regression method to conduct an empirical analysis of six factors that affect China’s grain production in 2000-2015. Area was a significant factor affecting grain output [22]. Considering the existence of multicollinearity, many scholars have also used ridge regression to reduce the impact of the collinearity problem on parameter estimation, making the regression results more economically significant. For example, Huang’s quantitative analysis of the factors affecting China’s food production based on revised CD production function and ridge regression and used the data from 1990 to 2008. The results showed that the increase in machinery input and improvement in irrigation conditions could increase food production, while labor input and chemical fertilizer input were not the driving factors for the increase of food production, and natural disasters still had a strong negative impact on food production, and food production showed increasing returns to scale [23].
From the perspective of food production forecasting, in addition to the above theories and models, time series models, support vector machines, and BP neural networks are the main research hotspots. Zhang et al. used data from 1978 to 2009 and the moving average model ARIMA to analyze and forecast China’s total food production. The results showed that the smooth ARIMA model has higher accuracy than the ARIMA model [24]. The filter analysis method separates China’s grain production from 1949 to 2008 into a time trend series and a fluctuation series. A polynomial model for the time series was established for the trend series, and a spectrum filtering method was used to estimate and fitted the fluctuation period of the grain output. The above two models were superimposed to predict China’s grain output in the next 10 years [25]; Li used the grey correlation analysis of the grain production system to determine the main impact factors in view of the complexity and incompleteness of the grain production system and combined support vector machines to build a prediction model [26]. In addition, many scholars have combined the grey correlation theory or stepwise regression method with BP neural network, the former determines the indicators used, and the latter is used to predict food production [27, 28]. The idea of the LASSO method is that the sum of absolute coefficients cannot be too large. Under this premise, applying the ordinary least squares method, the sum of squares of residuals is the least [28]. In order to ensure the dietary balance of Chinese residents, the influence factors of cereal consumption are valuable to research. They first use the LASSO method to select the main influence factors of cereal consumption, and then, they constructed a partially linear semiparametric model for predicting the cereal consumption of Chinese residents. The results show that the factors affecting per capita consumption of rice, wheat, and maize are different from one another and the three cereals have both common impact factors and differentiated ones; per capita disposable income is the common factor with a linear positive relationship to the consumption of the three cereals; the model constructed in this paper is well fitted and can accurately forecast the consumption of cereals; the average per capita consumption of rice, wheat, and maize is predicted to be 78.56 kg/year, 62.73 kg/year, and 6.64 kg/year, respectively, by 2025, which is excessive and is caused by irrational dietary structure, food wastage, and processing losses [29].

Other models use the whole information generated by spectral measurements, such as ridge regression, which was introduced by Tikhonov and generalized by Hoerl and Kennard [30, 31]. This type of multivariate linear regression includes a contraction of the multivariate model regression coefficients and reduces them to the same degree [25]. Hernandez et al.’s study evaluated the ability of canopy reflectance spectroscopy at the range from 350 to 2500 nm to predict grain yield in a large panel (368 genotypes) of wheat (Triticum aestivum L.) through multivariate ridge regression models. Plants were treated under three water regimes in the Mediterranean conditions of central Chile: severe water stress; mild water stress; and full irrigation with mean grain yields of 1655, 4739, and 7967 kg/ha⁻¹, respectively. Models developed from reflectance data during anthesis and grain filling under all water regimes explained between 77% and 91% of the grain yield variability, with the highest values in severe water stress condition. When individual models were used to predict yield in the rest of the trials assessed, models fitted during anthesis under mild water stress performed best. Combined models using data from different water regimes and each phenological stage were used to predict grain yield, and the coefficients of determination ($R^2$) increased to 89.9% and 92.0% for anthesis and grain filling, respectively. The model generated during anthesis in mild water stress was the best at predicting yields when it was applied to other conditions. Comparisons against conventional reflectance indices were made, showing lower predictive abilities. It was concluded that a ridge regression model using a data set based on spectral reflectance at anthesis or grain filling represents an effective method to predict grain yield in genotypes under different water regimes [32].

Generally speaking, in the study of the main influencing factors of food production, stepwise regression can only reduce variables without really solving the parameter estimation bias under multicollinearity. Ridge regression has a certain subjectivity in variable selection, which is only suitable for reducing multicollinearity. In the study of food production forecasting methods, the combined model method and models related to machine learning are favored by researchers, but under the premise of a small amount of data, it is easy to produce small training errors and large generalization errors, and its results of multiperiod prediction are inferior to the multiple linear regression method. Therefore, after considering various theories and methods, this paper selects the relevant explanatory variables as fully as possible and uses full subset regression, LASSO regression, and ridge regression analysis to further empirically analyze the selected variables, so as to reduce the impact of collinearity and identify the main influencing factors.

### 3. Variable Selection

From the perspective of input and output, the main factors affecting China’s food production can be divided into four categories: (1) natural conditions, (2) labor input, (3) technology, and (4) policies and other factors [12, 20, 21, 27, 33–35].

The constraints of natural conditions are mainly reflected in three aspects: land, water, and climate. For food production, land is the object of labor and the means of production of people and is the “mother of wealth.” Its area and soil type will directly determine the output of food. China has a vast territory and a large north-south span. The geographical and environmental differences cause uneven spatial and temporal distribution of water resources, and the differences in climatic conditions are large. Water resources are relatively scarce in many places, and droughts and floods are endless, which greatly limits food production [21, 27, 34, 35].

The labor input and the technical level need to be considered together. On the one hand, labor input, that is, the actual amount of labor input in the production process, with the socioeconomic level and other factors unchanged, increasing labor input can bring about an increase in food production. On the one hand, the improvement of technology means the
improvement of labor productivity, the improvement of land utilization rate, or the improvement of resource economic efficiency, so when other factors remain unchanged, greater output is obtained with less labor input [22].

Based on the above analysis and the availability of data, this paper selected 10 variables that affect food production, which are the sown area, the affected area, and the affected area when considering natural conditions. The number of people employed in the primary industry when considering labor input, and effective irrigation area, fertilizer application, rural electricity use, agricultural plastic film use and pesticide use when considering technology. See Tables 1 and 2 for details.

All the research data are from the official website of the National Bureau of Statistics and Data Center, and the time span is from 1991 to 2018. In addition, in order to eliminate the impact of dimensionality and explain the final results, this study also performed logarithmic processing on all data. The descriptive statistics of the data are shown in Table 2.

### 4. Correlation Analysis

Figure 1 is the scatter plot of the variables selected for the current study. It can be seen from Figure 1 that there is a clear linear relationship between the grain output and the selected variables. In addition to the sown area of grain crops, there are also obvious linear relationships among other variables, indicating that there may be multiple collinearity problems when directly performing regression. It is worth noting that with the increase in the number of employed people in the primary industry, the output of food has shown a downward trend. This does not mean that the decrease in the number of employed people in the primary industry has led to an increase in food production. It means that the improvement of the agricultural industry technology has largely made up for the reduction of labor input, indicating that when analyzing the issue of food production, it is possible to temporarily ignore the explanatory variable of labor input [27]. Further, it can be seen that it is impact of affected area and disaster area which affects the production of the country and has different

---

**Table 1: Variables.**

| Variables | Variable meaning |
|-----------|------------------|
| $y$       | Grain output (10,000 tons) |
| $x_1$     | Employment in the primary industry (10,000 people) |
| $x_2$     | The sown area of food crops (thousand hectares) |
| $x_3$     | Total power of agricultural machinery (10,000 kilowatts) |
| $x_4$     | Effective irrigation area (thousand hectares) |
| $x_5$     | Pure fertilizer application amount (10,000 tons) |
| $x_6$     | Rural electricity consumption (billion kilowatt hours) |
| $x_7$     | Consumption of agricultural plastic film (ton) |
| $x_8$     | Amount of pesticide used (10,000 tons) |
| $x_9$     | Affected area (thousand hectares) |
| $x_{10}$  | Disaster area (thousand hectares) |

**Table 2: Descriptive statistics.**

| Variable | Mean     | Maximum   | Minimum   | Std dev.  |
|----------|----------|-----------|-----------|-----------|
| $y$      | 10.86270817 | 11.09984221 | 10.67057107 | 0.14594216 |
| $x_1$    | 10.32984707 | 10.57382659 | 9.91630506 | 0.20978288 |
| $x_2$    | 11.61474   | 11.68881   | 11.50701   | 0.04777608 |
| $x_3$    | 11.05192624 | 11.62382325 | 10.28836212 | 0.43968355 |
| $x_4$    | 10.93997146 | 11.13124973 | 10.77524253 | 0.11156401 |
| $x_5$    | 8.43068222  | 8.70327434  | 7.93919447  | 0.23065250 |
| $x_6$    | 8.2265766   | 9.1616143   | 6.8702611   | 0.7371465  |
| $x_7$    | 14.27251711 | 14.77239052 | 13.37256941 | 0.43000708 |
| $x_8$    | 4.92201629  | 5.19722550  | 4.33768282  | 0.25639565 |
| $x_9$    | 10.58050597 | 10.92363367 | 9.82433611  | 0.33470965 |
| $x_{10}$ | 9.91449979  | 10.44505574 | 9.12706745  | 0.38274362 |
impact on the country’s selected variables. And it can also be seen that except affected area and disaster areas, all other variables are having direct relationship between each other. The sown area of food crops shows seasonal relationship to all the selected variables and has seasonal ups and downs.

Figure 2 is the correlation coefficient matrix of the variables. From Figure 2, we can see that (1) the absolute values of the correlation coefficients between the selected independent variable and the dependent variable are larger than 0.7, indicating that the linear relationship is obvious. (2) In addition to the sown area of grain crops, there are also obvious linear correlations between other independent variables, indicating that there may be serious multicollinearity problems when directly performing regression [28].

Regarding the testing of collinearity problems, in addition to using correlation coefficients to assist judgment, there are currently two main methods: (1) judging whether there is a collinearity problem based on the condition number and (2) judging whether a collinearity problem exists based on the variance inflation factor [22].

The mathematical expression of the condition number is as follows:

$$k = \frac{\lambda_{\text{max}}(X^TX)}{\lambda_{\text{min}}(X^TX)},$$  \hspace{1cm} (1)
where $\lambda$ is the eigenvalue and $X$ is the independent variable matrix. Generally speaking, when $k < 100$, the degree of multicollinearity is considered small; when $100 \leq k \leq 1000$, multicollinearity is considered to exist; when $k > 1000$, serious multicollinearity is considered to exist. After removing the independent variable $x_1$, the condition number of the independent variable matrix composed of the remaining independent variables is 3984.441, which is much greater than 1,000 [7, 8, 34].

The formula for calculating the variance inflation factor is as follows:

$$VIF_i = \frac{1}{1 - R_i^2}.$$  \hspace{1cm} (2)

$R_i^2$ represents $R^2$ obtained by linear regression using variable $x_i$ as the dependent variable and the other $k - 1$ variables as independent variables. Generally, we think that there is a multicollinearity problem when the variance inflation factor is greater than 5 or 10. When the variance inflation factor is greater than 100, there is a serious multicollinearity problem [3, 7, 25].

After calculation, the variance inflation factor of various variables is shown in Table 3. From this table, we can clearly see that except for $x_2$, the variance inflation factor of other variables is all higher than 10, and most of them are higher than 100, which indicates that there is a serious multicollinearity problem between independent variables, and linear regression cannot be performed directly.

5. Empirical Analysis

5.1. Full Subset Regression Analysis. Stepwise regression means that the model will add (or drop) a variable one by one until it reaches a certain stopping criterion. However, in practice, although stepwise regression analysis can find a good model, it cannot guarantee that the model is the best one. Therefore, this paper uses full subset regression analysis to analyze the possible combination of variables [34, 35]. The results are shown in Figure 3.

From Figure 3, we can see that based on the adjusted $R$-squared, the optimal subset contains five variables, which are $x_3, x_5, x_6, x_7$, and $x_{10}$. Based on this, a regression equation is established, and the results are shown in Table 4.

From the table above, we can see that the regression coefficients of the five explanatory variables are all significantly nonzero, but the regression coefficient of the total power of the agricultural machinery represented by the $x_3$ variable is negative [25, 28]. The main reason for this is the existence of multicollinearity [13, 16].

5.2. LASSO Regression Analysis. LASSO regression is a biased estimation regression method that can be used for collinear data analysis. It differs from ordinary least squares regression in that it adds a 1-norm penalty term in the estimation process [14, 28, 29], which is expressed as a mathematical formula as follows:

$$\hat{\theta} = \arg\min_\theta ||X\theta - y||_2^2 + C||\theta||_1.$$  \hspace{1cm} (3)

Using R 3.6.2 for LASSO regression, the Cp value calculation results are shown in the following table.

Table 5 shows the value of the Cp statistic at each step of variable selection. Among them, it can be seen that the Cp value of step 10 is the smallest, which is 4.6649. The corresponding coefficients of each variable in Table 6 can be seen, only $x_3, x_5, x_6, x_9$, and $x_{10}$ are not zero, and the coefficients are 1.420427, 0.196403, 0.043928, -0.03694, and -0.0474. It shows that the sown area of food crop has a significant impact on food production, and pure fertilizer application amount, rural electricity consumption, and disaster area can affect grain production as well.

5.3. Ridge Regression Analysis. Ridge regression is also a biased estimation regression method that can be used for collinear data analysis. The regression coefficient is more practical and reliable at the cost of losing some information and reducing accuracy, and it fits a large number of conditions (morbid data) [31, 32]. Other models use the whole information generated by spectral measurements, such as ridge regression, which was introduced by Li and Ren and generalized by Li et al. [27, 28]. This type of multivariate linear regression includes a
Table 4: Full subset regression results.

| Variable | Coefficients | Std error | T value | P value |
|----------|--------------|-----------|---------|---------|
| Constant | -6.93797     | 0.84836   | -8.178  | 4.09e-08*** |
| $x_2$    | 1.44285      | 0.06502   | 22.190  | <2e-16*** |
| $x_3$    | -0.11074     | 0.04638   | -2.388  | 0.02597*  |
| $x_5$    | 0.23684      | 0.07007   | 3.380   | 0.00270** |
| $x_6$    | 0.10630      | 0.02999   | 3.545   | 0.00182** |
| $x_{10}$ | -0.06103     | 0.01279   | -4.770  | 9.20e-05*** |

$R^2$ 0.9948  
Adjusted $R^2$ 0.9936

| $F$ value | $P$ value |
|-----------|-----------|
| 839.1     | <2.2e-16  |

DW 2.2278

Table 5: Changes in Cp values of the data in LASSO regression.

| Step | Df | Rss   | Cp    |
|------|----|-------|-------|
| 0    | 1  | 0.57508 | 3662.8281 |
| 1    | 2  | 0.44547 | 2833.4618 |
| 2    | 3  | 0.27043 | 1712.7045 |
| 3    | 4  | 0.24992 | 1583.0999 |
| 4    | 5  | 0.09332 | 580.5712  |
| 5    | 4  | 0.04680 | 280.2289  |
| 6    | 5  | 0.01993 | 109.8113  |
| 7    | 6  | 0.00783 | 34.2075   |
| 8    | 7  | 0.00404 | 11.8952   |
| 9    | 6  | 0.00362 | 7.2435    |
| 10   | 7  | 0.00291 | 4.6649    |
| 11   | 8  | 0.00289 | 6.5284    |
| 12   | 9  | 0.00286 | 8.3747    |
| 13   | 10 | 0.00281 | 10.0000   |

Table 6: LASSO regression coefficients at each step.

| Step | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| 1    | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0        |
| 2    | 0     | 0     | 0.171949 | 0     | 0     | 0     | 0     | 0     | 0        |
| 3    | 0     | 0     | 0.316865 | 0     | 0     | 0     | 0     | 0     | -0.0483  |
| 4    | 0     | 0     | 0.339717 | 0     | 0     | 0     | 0     | 0     | -0.0577  |
| 5    | 0.45343 | 0     | 0.554411 | 0     | 0     | 0     | 0     | 0     | -0.04692 |
| 6    | 0.672158 | 0     | 0.639493 | 0     | 0     | 0     | 0     | 0     | -0.05293 |
| 7    | 0.938281 | 0     | 0.445622 | 0.112526 | 0     | 0     | 0     | 0     | -0.0617  |
| 8    | 1.153794 | 0     | 0.248814 | 0.211913 | 0     | 0     | 0     | 0     | -0.02147 |
| 9    | 1.367226 | 0     | 1.194545 | 0.043928 | 0     | 0     | 0     | 0     | -0.03669 |
| 10   | 1.420427 | 0     | 0.196403 | 0.047311 | 0     | 0     | 0     | 0     | -0.03094 |
| 11   | 1.422217 | -0.08757 | 0     | 0.243953 | 0.086609 | 0     | 0     | 0     | -0.02709 |
| 12   | 1.42105 | -0.09139 | 0     | 0.262388 | 0.086814 | 0     | -0.00946 | -0.02559 | -0.04645 |
| 13   | 1.42738 | -0.0914 | -0.01892 | 0.284621 | 0.087443 | 0     | -0.02325 | -0.02566 | -0.04562 |
| 14   | 1.473969 | -0.09244 | -0.1296 | 0.371263 | 0.089772 | 0.018195 | -0.09128 | -0.02441 | -0.04353 |
contraction of the multivariate model regression coefficients and reduces them to the same degree [25].

Compared with LASSO regression, the ridge regression adds a 2-norm penalty term in the estimation process, which is expressed as a mathematical formula as follows:

$$\hat{\theta} = \arg\min_{\theta} \|X\theta - y\|_2^2 + C\|\theta\|_2^2.$$  \hspace{1cm} (4)

Using R 3.6.2 for direct ridge regression, the results are shown in Table 7:

| Variable | Regression coefficients | Standardized regression coefficients | Standard error | T value | P value   |
|----------|-------------------------|--------------------------------------|----------------|---------|-----------|
| Intercept | -6.022672               | 0.319030                             | 0.017076       | 18.683  | <2e-16*** |
| $x_2$    | 1.285106                | 0.005408                             | 0.012356       | 0.021791 | 0.567     | 0.570710  |
| $x_3$    | 0.125080                | 0.083995                             | 0.100669       | 0.03788 | 3.048     | 0.002303**|
| $x_4$    | 0.083995                | 0.023191                             | 0.088829       | 0.020001 | 4.441     | 0.894e-06***|
| $x_5$    | 0.005408                | 0.022444                             | 0.050148       | 0.017251 | 2.907     | 0.003650**|
| $x_6$    | 0.059953                | 0.043730                             | 0.079874       | 0.022066 | 3.620     | 0.000295**|
| $x_7$    | 0.02624                 | -0.043730                            | -0.076056      | 0.025652 | 2.965     | 0.003028**|
| $x_8$    | 0.023191                | -0.052508                            | -0.104428      | 0.024886 | 4.196     | 2.71e-05***|

Compared with LASSO regression, the ridge regression adds a 2-norm penalty term in the estimation process, which is expressed as a mathematical formula as follows:

$$\hat{\theta} = \arg\min_{\theta} \|X\theta - y\|_2^2 + C\|\theta\|_2^2.$$  \hspace{1cm} (4)

Using R 3.6.2 for direct ridge regression, the results are shown in Table 7:

It can be seen from the above table that the variable $x_3$'s coefficient is positive [27], which is in line with economic laws, but the coefficient is not significantly nonzero. Therefore, the $x_3$ variable is excluded and we perform ridge regression again. The final ridge regression result is shown in Table 8.

The ridge parameter after removing the $x_3$ variable is 0.03450807. It is relatively small, which indicates that although the estimation is biased, the deviation from the least square estimation result is not large. Comparing the standardized regression coefficients, we can clearly see that the sown area of food crops has a greater positive impact on food production, followed closely by the application of pure agricultural fertilizers and rural power consumption, and the negative impact of disaster area on food production is relatively large, which is basically consistent with LASSO regression results.

Table 8: Final ridge regression results.

| Variable | Regression coefficients | Standardized regression coefficients | Standard error | T value | P value   |
|----------|-------------------------|--------------------------------------|----------------|---------|-----------|
| Intercept | -6.26961                | 0.32595                              | 0.01721        | 18.941  | <2e-16*** |
| $x_2$    | 1.31299                 | 0.11369                              | 0.06591        | 0.02834 | 2.326     | 0.020025* |
| $x_4$    | 0.09098                 | 0.02624                              | 0.10904        | 0.01537 | 7.093     | 1.31e-12***|
| $x_5$    | 0.02254                 | 0.05997                              | 0.07989        | 0.02434 | 3.282     | 0.001030**|
| $x_7$    | 0.05997                 | -0.04163                             | -0.07241       | 0.02957 | 2.448     | 0.014348* |
| $x_8$    | -0.02624                | -0.05252                             | -0.10445       | 0.02747 | 3.803     | 0.000143***|

6. Conclusion

Considering the results of the previous regression analysis, we can draw the following conclusions:

(1) The increase in the sown area of food crops is the main reason for the increase in food production. After comprehensively analyzing of the regression results of the three regressions, we find that the elasticity coefficient of the variable $x_3$ is all above 1, indicating that the increase in the sown area can significantly increase the growth of grain output.

(2) The use of agricultural fertilizers and the increase in rural electricity consumption are the driving factors for the increase in grain output. From the empirical analysis, we can see that, in addition to the sown area, the amount of fertilizer applied to agriculture and rural electricity consumption are significantly positive in the three models, and the regression coefficient is relatively large than other variables, indicating that they have indeed effectively promoted the increase in food production.

(3) The total power of agricultural machinery has a limited impact on grain output. The effects of the total...
power of agricultural machinery on grain yield are not the same in the three models. The elasticity coefficient of the total power of agricultural machinery is negative in the full subset regression, and the elastic coefficient is positive in the ridge regression and LASSO regression but not significant. The main reason for this result significance might be due to the suitable variables selecting procedure listed in Figure 3 where to run the full subset regression only 5 variables are selected to show their impact on the dependent variable. Second, it might be the time span of the data is too long, which causes the effect of mechanical inputs on grain output to be not obvious.

(4) Natural disasters have a certain negative impact on food production. From the above three regression models, we can see that for each 1% increase in the disaster area, food production will decrease by about 0.05%, indicating that natural disasters still have a certain negative impact on food production. It is still necessary to consolidate the ability of agricultural production to resist natural disasters.

Data Availability
All data, models, and code generated or used during the study appear in the submitted article.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Acknowledgments
This study was supported by the Fundamental Research Funds for the Central Universities (Grant Nos. HIT.HSS.202233, HIT.HSS.202118, and HIT.HSS.201851).

References
[1] Z. Xiaoshi, L. Gucheng, L. Cheng, and H. Capital, “Land scale and agricultural production efficiency,” Journal of HZAU(Social Sciences Edition), vol. 2, pp. 8–18, 2018.
[2] Y. Wei, Z. Peng, and J. Zhiheng, “Spatial interaction effect of rural education human capital and agricultural total factor productivity in China—empirical analysis based on spatial simultaneous equations,” Journal of China Agricultural University, vol. 25, no. 3, pp. 192–202, 2020.
[3] L. Jingbo and G. Yuan, “Research on the influence mechanism of China’s population aging on labor productivity,” Nankai Economic Studies, vol. 3, pp. 61–80, 2020.
[4] W. Yuqi and L. Lei, “The influence of new rural cooperative medical insurance and health human capital on labor participation of rural residents,” Chinese Rural Economy, vol. 11, pp. 68–81, 2016.
[5] Y. Qiuxia, C. Zhaoju, and X. Huizing, “Labor allocation efficiency of farmer households and its influencing factors: based on the survey data of 637 farmer households in Jiangxi province,” Journal of Hunan Agricultural University(Social Science), vol. 19, no. 5, pp. 11–18, 2018.
[6] R. Cerda, C. Allinne, C. Gary et al., “Effects of shade, altitude and management on multiple ecosystem services in coffee agroecosystems,” European Journal of Agronomy, vol. 82, pp. 308–319, 2017.
[7] M. H. Ahmed and K. A. Melesse, “Impact of off-farm activities on technical efficiency: evidence from maize producers of eastern Ethiopia,” Agricultural and Food Economics, vol. 6, no. 1, pp. 1–20, 2018.
[8] S. Han, D. Li, and J. Xiong, “Changes in the amount of cultivated land resources in Guangzhou and their impact on food security,” Journal of Agricultural and Forestry Economic Management, vol. 6, 2016.
[9] Z. Feng, “Food security and cultivated land security for China’s future population development,” Population Research, vol. 2, pp. 17–31, 2007.
[10] T. Taotao, Q. Ma, and G. Li, “Population aging, population urbanization and China’s food security: simulation based on China CGE model,” Journal of Zhongnan University of Economics and Law, vol. 4, 2017.
[11] J. Xiang and F. Zhong, “Impact of demographic transition on food demand in China: 2010–2050,” China Population, Resources and Environment, vol. 23, no. 6, pp. 117–121, 2013.
[12] S. M. Sanusi and I. P. Singh, “Empirical analysis of economies of scale and cost efficiency of small-scale maize production in Niger state, Nigeria,” Indian Journal of Economics and Development, vol. 12, no. 1, pp. 55–64, 2016.
[13] W. Xuejiao and X. Haiying, “Spatial correlation measurement and influence factors of allocation efficiency for maize production in China,” Social Science Journal of Harbin Institute of Technology, vol. 18, no. 6, pp. 125–131, 2016.
[14] L. Wang, “Research on the impact of China’s population urbanization on food security,” Lanzhou University, China, 2019.
[15] X. Cai, “Research on the impact of urbanization development in Henan Province on food security,” Henan Agriculture, vol. 11, pp. 8–9 + 64, 2015.
[16] X. Du, “Empirical analysis of influencing factors of food security in China,” Anhui Agricultural Science, vol. 43, no. 5, 2015.
[17] W. Chen and X. Xie, “Research on the impact mechanism of climate disasters on food security,” Issues in Agricultural Economy, vol. 34, no. 1, pp. 12–19, 2013.
[18] C. Chih-Ming, “Interpretive structural modeling apply to patient safety culture,” International Journal of Medicine and Pharmaceutical Sciences, vol. 8, no. 4, pp. 21–26, 2018.
[19] J. Qiu, S. Wang, and G. Gao, “Global warming and food security in China—a study based on GTAP model,” Finance Science, no. 4, pp. 57–67, 2015.
[20] S. Yin, L. Wu, and Y. Zhang, “Empirical analysis of the factors affecting the grain output fluctuation in China,” Systems Engineering Theory & Practice, vol. 10, pp. 28–34, 2009.
[21] S. Zai, S. Guo, and J. Wen, “Analysis of influencing factors of grain production in Henan Province,” People’s Yellow River, vol. 33, no. 12, pp. 94–96128, 2011.
[22] W. Qiling and K. Zhang, “Analysis of influencing factors of grain output based on regression analysis,” Grain Science and Technology and Economy, vol. 42, no. 6, 2017.
[23] Z. Huang, “Analysis of influencing factors of China’s grain production based on ridge regression analysis of C-D.
production function,” *Taxation and Economy*, vol. 5, pp. 50–54, 2014.

[24] H. Zhang, F. Gao, and X. Jin, “Time series analysis of grain yield in China based on ARIMA model,” *Science and Technology Information*, vol. 30, no. 30, pp. 533–534, 2010.

[25] D. Sun and S. Liang, “Research on time series model and application of grain yield forecast in China,” *Agricultural Technology and Economy*, vol. 3, pp. 99–108, 2010.

[26] X. Li, “Grain output forecasting model in China based on grey relational support vector machine,” *Journal of Hebei Polytechnic University (Natural Science Edition)*, vol. 4, pp. 81–85 + 108, 2008.

[27] H. Li and Z. Ren, “Grey correlation analysis of grain yield forecast and influencing factors in Shanxi Province,” *Research on Land and Natural Resources*, vol. 4, pp. 58–60, 2009.

[28] P. Li, X. Yan, and X. Dongrui, “Comparison of BP neural network and multiple linear regression in prediction of spatial distribution of grain yield,” *Journal of Arid Land Resources and Environment*, vol. 28, no. 9, pp. 74–79, 2014.

[29] Y. Ma, “A new forecasting method for cereal consumption: combined method of LASSO and semi-parametric regression,” in *AIP Conference Proceedings*, Wuhan, China, 2019.

[30] A. Tikhonov, “On the stability of inverse problems,” *Proceedings of the USSR Academy of Sciences*, vol. 39, pp. 195–198, 1943.

[31] A. E. Hoerl and R. W. Kennard, “Ridge regression: applications to nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 69–82, 1970.

[32] J. Hernandez, G. Lobos, I. Matus, A. del Pozo, P. Silva, and M. Galleguillos, “Using ridge regression models to estimate grain yield from field spectral data in bread wheat (*Triticum aestivum* L.) grown under three water regimes,” *Remote Sensing*, vol. 7, no. 2, pp. 2109–2126, 2015.

[33] F. Zeqiang, Y. Cai, and Y. Yang, “Correlation analysis between China’s food security and cultivated land resources change,” *Journal of Natural Resources*, vol. 4, pp. 17–23, 2001.

[34] G. A. Blackburn, “Hyperspectral remote sensing of plant pigments,” *Journal of Experimental Botany*, vol. 58, no. 4, pp. 855–867, 2007.

[35] D. C. Lazaridis, J. Verbesselt, and A. P. Robinson, “Penalized regression techniques for prediction: a case study for predicting tree mortality using remotely sensed vegetation indices this article is one of a selection of papers from extending forest inventory and monitoring over space and time,” *Canadian Journal of Forest Research*, vol. 41, pp. 24–34, 2011.