ISSN: 2688-9323

Analysis of Factors Affecting Grain Production in China

Xinyu Xie

Institute of Finance and Public Management, Anhui University of Finance and Economics, Bengbu, Anhui 233030, China

Abstract

This study focuses on the influencing factors of China's grain yield and utilizes econometric methods for in-depth analysis. A suitable econometric model is constructed by collecting data from various aspects, including total power of agricultural machinery, effective irrigated area, pure amount of agricultural fertilizer applied, sown area of grain crops and affected area indicators. The empirical analysis of the degree and direction of the influence of each factor on China's grain production shows that some factors, such as effective irrigated area and the amount of agricultural fertilizer used, have a significant positive correlation with grain production. The conclusions of the study help to understand the internal mechanism of China's grain production system, and provide valuable theoretical basis and data support for the formulation of scientific and reasonable agricultural policies to guarantee food security and increase grain production.

Keywords

Agriculture; Food Production; Effective Irrigated.

1. Background to the Study and Formulation of Questions

With global population growth and climate change, food security has become a focus of international attention, and food self-sufficiency is crucial given our large population base. Economic development has led to an upgrading of the population's dietary structure, which has increased the demand for food from quantity to quality, while the modernization of agriculture has resulted in a shift in agricultural production methods, with new technologies and inputs affecting food production. However, resource scarcity and environmental problems, such as the reduction of arable land, uneven water resources and pollution, have constrained the increase in food production. Against this complex background, analyzing the factors affecting China's food production is of great significance in guaranteeing food security, promoting sustainable agricultural development and meeting social needs.

Food security, as an important foundation for national security, is a matter of national livelihood and social stability. In China, as the population continues to grow, the living standards of the population continue to improve and the economy develops rapidly, the demand for food is growing rigidly, and it is of vital strategic significance to ensure stable growth in food production. Grain production is affected by the interaction of many complex factors, and clarifying the mechanism of these factors on grain production is crucial for the formulation of accurate and effective agricultural policies, rational allocation of agricultural resources and guaranteeing the long-term stable supply of grain. The purpose of this study is to use econometric methods to systematically and deeply analyze the influencing factors of food production in China.

For this purpose, key indicators such as total power of agricultural machinery, effective irrigated area, net application of agricultural fertilizers, sown area of food crops and affected area were selected. Has the increase in the total power of agricultural machinery actually contributed to the increase in food production? To what extent has the expansion of the

ISSN: 2688-9323

effective irrigated area ensured the stability of food production? How do changes in the discounted amount of agricultural fertilizers applied affect food production, and are there diminishing marginal benefits? What is the direct correlation between fluctuations in the area sown to food crops and food production? To what extent does the size of the affected area impact on grain yields? Through the construction of scientific and reasonable econometric models, these questions are analyzed precisely and quantitatively, in order to provide a solid theoretical basis and practical guidance for the sustainable development of China's grain industry, so as to better grasp the laws of grain production under the complex and volatile environment, improve the efficiency of grain production and yield level, and stabilize the defense line of national food security.

2. Data Collection

By checking the annual data about agriculture from the National Bureau of Statistics, we got the data about the total power of agricultural machinery, effective irrigated area, discounted amount of agricultural fertilizer application, sown area of grain crops and affected area from 2004 to 2023, which are collated and shown in the table below.

Table 1. Data table on food production and related inputs in China, 2004-2023

particular year	Grain production (tons)	Total power of agricultural machinery (10,000 kilowatts)	Effective irrigated area (thousands of hectares)	Agricultural fertilizer application (tons)	Area sown with food crops (thousands of hectares)	Area affected (thousands of hectares)
	Y	X1	X2	X3	X4	X5
2004	46946.95	64027.91	54478.42	4636.58	101606.03	37110
2005	48402.19	68397.85	55029.34	4766.22	104278.38	38820
2006	49804.23	72522.12	55750.5	4927.69	104957.7	41090
2007	50413.85	76589.56	56518.34	5107.83	105998.62	48990
2008	53434.29	82190.41	58471.68	5239.02	107544.51	39990
2009	53940.86	87496.1	59261.45	5404.35	110255.09	47210
2010	55911.31	92780.48	60347.7	5561.68	111695.42	37430
2011	58849.33	97734.66	61681.56	5704.24	112980.35	32470
2012	61222.62	102558.96	62490.52	5838.85	114368.04	24960
2013	63048.2	103906.75	63473.3	5911.86	115907.54	31350
2014	63964.83	108056.58	64539.53	5995.94	117455.18	24890
2015	66060.27	111728.07	65872.64	6022.6	118962.81	21770
2016	66043.51	97245.59	67140.62	5984.41	119230.06	26220
2017	66160.73	98783.35	67815.57	5859.41	117989.06	18480
2018	65789.22	100371.74	68271.64	5653.42	117038.21	20814
2019	66384.34	102758.26	68678.61	5403.59	116063.6	19257
2020	66949.15	105622.15	69160.52	5250.65	116768.17	19958
2021	68284.75	107764.32	69609.48	5191.26	117630.82	11739
2022	68652.77	110597.19	70358.87	5079.2	118332.11	12072
2023	69540.99	113742.57	71644	5021.74	118968.54	10539

(Source: National Statistical Office - annual data).

ISSN: 2688-9323

3. Modeling

3.1. Scatterplotting

Y, X1, X2, X3, X4, X5, respectively, denote grain production, total power of agricultural machinery, effective irrigated area, discounted agricultural fertilizer application, sown area of grain crops and affected area. Then scatter plots of the explanatory variables (Y) and explanatory variables (X1, X2, X3, X4) were plotted separately and the results are shown below:

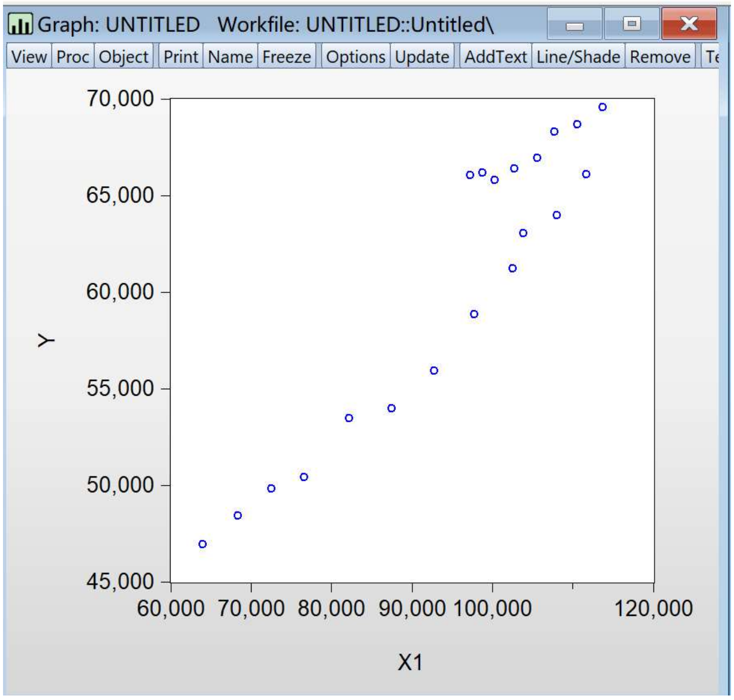


Figure 1. Scatter plot of Y and X1

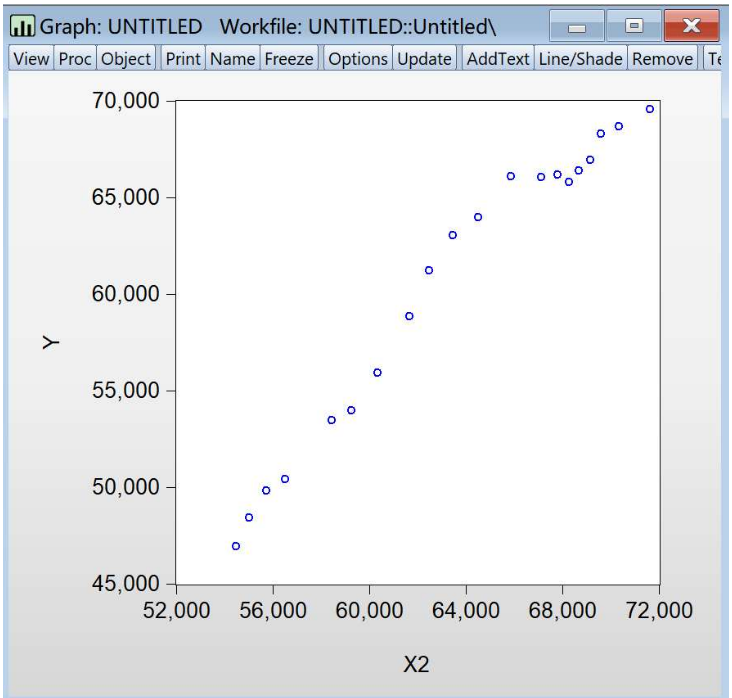


Figure 2. Scatter plot of Y and X2

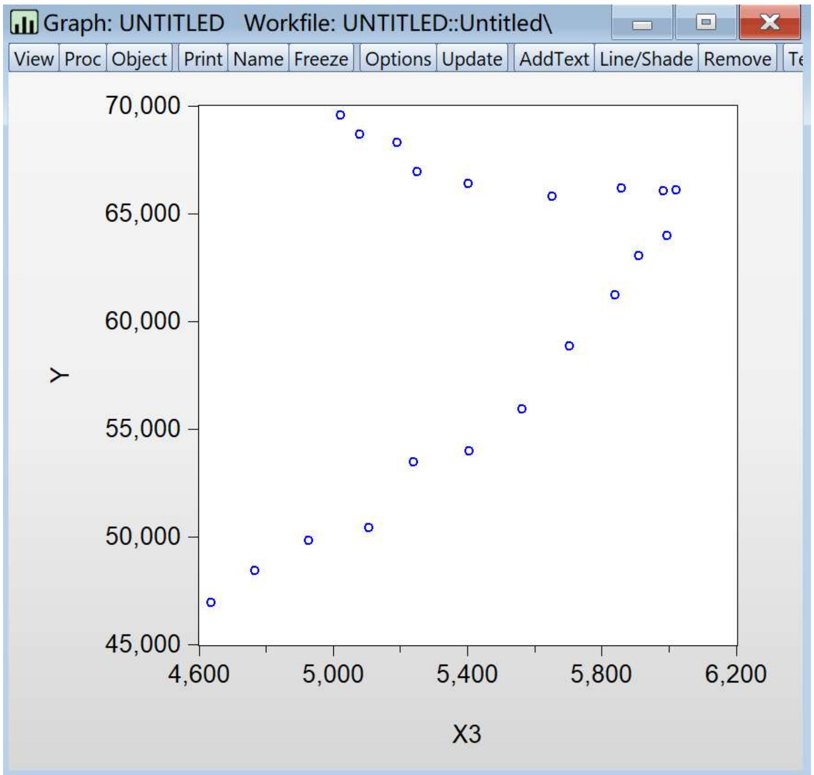


Figure 3. Scatter plot of Y and X3

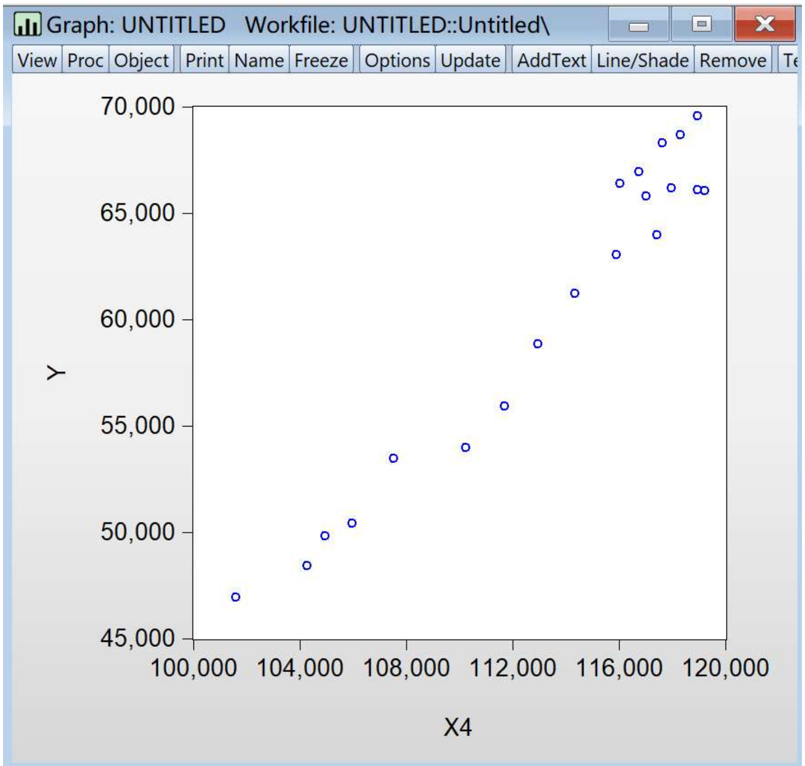


Figure 4. Scatter plot of Y and X4

ISSN: 2688-9323

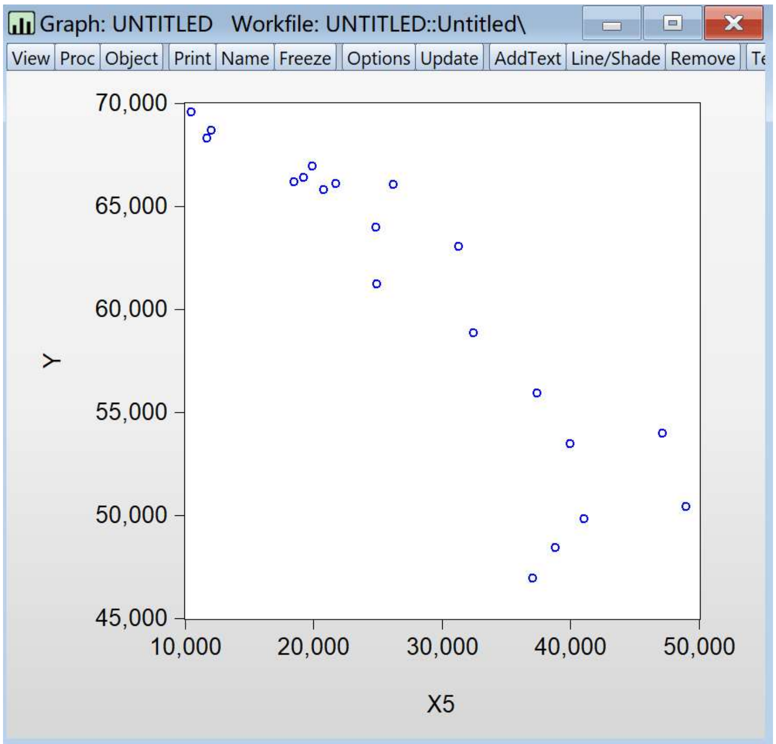


Figure 5. Scatter plot of Y and X5

3.2. Establishment of Multiple Regression Models

According to the results of the scatter plot, the linear relationship between the explanatory variables and the explanatory variables is not clear, according to the mathematical principle, taking the logarithm can eliminate the influence brought by the index, so that the variables maintain a linear relationship, and so a new regression equation is established:

$$LNY=\beta_0+\beta_1 \ln X_1+\beta_2 \ln X_2+\beta_3 \ln X_3+\beta_4 \ln X_4+\beta_5 \ln X_5+u$$

3.3. OLS Method of Estimating Models

Using EViews 9.0, after creating the working file and inputting the data, the variables are redefined and then the OLS method is applied to estimate the model parameters. The command approach is as follows:

DATA Y X1 X2 X3 X4 X5
GENR LNY=LOG(Y)
GENR LNX1=LOG(X1)
GENR LNX2=LOG(X2)
GENR LNX3=LOG(X3)
GENR LNX4=LOG(X4)
GENR LNX5=LOG(X5)
LS LNY C LNX1 LNX2 LNX3 LNX4 LNX5

The results of the model estimation are as follows:

$$LNY=-4.7143+0.0914\ln X1+0.8579\ln X2+0.2141\ln X3+0.3195\ln X4+0.0367\ln X5$$

$$s= (0.0467)\ (0.1547)\ (0.0908)\ (0.3881)\ (0.0152)$$

$$t=(1.9586)\ (5.5470)\ (2.3563)\ (0.8234)\ (-2.421)$$

$$R2 =0.9962 \quad \bar{R}(2) =0.9949 \quad F=739.5219 \quad n=20$$

ISSN: 2688-9323

It can be seen from the above regression results that the values of R2 and $\bar{R}(2)$ are closer to 1, which indicates that the model is well fitted.The F-statistic of 739.5219 is significant for the model. The t-test of the coefficients of the explanatory variables lnX2, lnX3, and lnX5 are significant at the significance level of α=0.05, and the t-test of the coefficients of lnX1 and lnX4 are not significant. This indicates that there may be multicollinearity, so further multicollinearity tests are required.

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 15:11
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.714324	2.445981	-1.927376	0.0745
LNX1	0.091396	0.046664	1.958585	0.0704
LNX2	0.857922	0.154665	5.546969	0.0001
LNX3	0.214056	0.090845	2.356286	0.0336
LNX4	0.319514	0.388055	0.823373	0.4241
LNX5	-0.036735	0.015176	-2.420651	0.0297

R-squared	0.996228	Mean dependent var	11.00237
Adjusted R-squared	0.994881	S.D. dependent var	0.130543
S.E. of regression	0.009340	Akaike info criterion	-6.265689
Sum squared resid	0.001221	Schwarz criterion	-5.966969
Log likelihood	68.65689	Hannan-Quinn criter.	-6.207375
F-statistic	739.5219	Durbin-Watson stat	1.439405
Prob(F-statistic)	0.000000		

Figure 6. The results of OLS method model estimation of double logarithm model

4. Testing and Adjustment of the Model

4.1. Economic Significance Test

From the economic analysis, it can be seen that grain production (Y) should be positively correlated with the total power of agricultural machinery (X1), the effective irrigated area (X2), the discounted amount of agricultural fertilizers applied (X3), and the sown area of grain crops (X4), and negatively correlated with the area affected by disasters (X5).

4.2. Tests and Corrections for Multicollinearity

4.2.1. Variance Inflation Factor Test

In EViews, you can directly calculate the variance inflation factor of the explanatory variables, click View/Coefficient Diagnostics/Variance Inflation Factors in Equation regression results, where Centered VIF is the variance inflation factor, and the results are shown below.

If the variance inflation factor VIF≥ 10, it usually indicates that there is a serious multicollinearity between that explanatory variable and the rest of the explanatory variables, and here all five explanatory variables have a variance inflation factor greater than 10, which suggests that there is a serious multicollinearity problem.

ISSN: 2688-9323

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Variance Inflation Factors
Date: 12/25/24 Time: 17:15
Sample: 2004 2023
Included observations: 20

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	5.982821	1371637.	NA
LNX1	0.002178	65475.50	13.98946
LNX2	0.023921	670352.0	41.34919
LNX3	0.008253	139826.3	11.62274
LNX4	0.150587	4675676.	85.37685
LNX5	0.000230	5456.592	10.69746

Figure 7. Results of variance inflation factor test

4.2.2. Correction of Multicollinearity Using Stepwise Regression Method

(1) Using EViews software command (COR lnY lnX1 lnX2 lnX3 lnX4 lnX5) to get the correlation coefficient matrix between the explained variable lnY and the explanatory variables (see the figure below), it is found that lnX2 is the explanatory variable that has the closest relationship with the explanatory variables, and in this way, the one-way regression equation can be established.

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Correlation

	LNY	LNX1	LNX2	LNX3	LNX4	LNX5
LNY	1.000000	0.952470	0.983958	0.548822	0.981520	-0.851694
LNX1	0.952470	1.000000	0.907379	0.634867	0.962826	-0.746665
LNX2	0.983958	0.907379	1.000000	0.418270	0.944766	-0.894505
LNX3	0.548822	0.634867	0.418270	1.000000	0.668048	-0.094768
LNX4	0.981520	0.962826	0.944766	0.668048	1.000000	-0.766768
LNX5	-0.851694	-0.746665	-0.894505	-0.094768	-0.766768	1.000000

Figure 8. A matrix of correlation coefficients between explained variables and explanatory variables

Enter "LS lnY C lnX2" in the command window and the result is shown below:

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 17:59
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.937904	0.681227	-7.248545	0.0000
LNX2	1.441837	0.061617	23.40008	0.0000

R-squared	0.968173	Mean dependent var	11.00237
Adjusted R-squared	0.966405	S.D. dependent var	0.130543
S.E. of regression	0.023927	Akaike info criterion	-4.532976
Sum squared resid	0.010305	Schwarz criterion	-4.433403
Log likelihood	47.32976	Hannan-Quinn criter.	-4.513538
F-statistic	547.5636	Durbin-Watson stat	0.310817
Prob(F-statistic)	0.000000		

Figure 9. The test results were modified after multicollinearity

(2) Introduce lnX1, lnX3, lnX4, lnX5 into the univariate regression equation and estimate the four bivariate regression equations respectively (the results are shown below). Then economic significance test, t-test and comparison $\overline{R}(2)$ were conducted to select the most appropriate regression equation as lnY=f(lnX2, lnX4).

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 18:03
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.913483	0.565994	-5.147554	0.0001
LNX2	0.992912	0.091412	10.86195	0.0000
LNX1	0.256633	0.047416	5.412311	0.0000

R-squared	0.988312	Mean dependent var	11.00237
Adjusted R-squared	0.986937	S.D. dependent var	0.130543
S.E. of regression	0.014920	Akaike info criterion	-5.434756
Sum squared resid	0.003784	Schwarz criterion	-5.285396
Log likelihood	57.34756	Hannan-Quinn criter.	-5.405599
F-statistic	718.7694	Durbin-Watson stat	0.665142
Prob(F-statistic)	0.000000		

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 18:05
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.132285	0.414545	-14.79282	0.0000
LNX2	1.339869	0.037100	36.11494	0.0000
LNX3	0.270081	0.041102	6.571012	0.0000

R-squared	0.991009	Mean dependent var	11.00237
Adjusted R-squared	0.989951	S.D. dependent var	0.130543
S.E. of regression	0.013086	Akaike info criterion	-5.697073
Sum squared resid	0.002911	Schwarz criterion	-5.547713
Log likelihood	59.97073	Hannan-Quinn criter.	-5.667916
F-statistic	936.9055	Durbin-Watson stat	0.850902
Prob(F-statistic)	0.000000		

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 18:06
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11.93075	0.936543	-12.73914	0.0000
LNX2	0.772819	0.089034	8.680012	0.0000
LNX4	1.236453	0.155461	7.953444	0.0000

R-squared	0.993259	Mean dependent var	11.00237
Adjusted R-squared	0.992465	S.D. dependent var	0.130543
S.E. of regression	0.011331	Akaike info criterion	-5.985000
Sum squared resid	0.002183	Schwarz criterion	-5.835640
Log likelihood	62.85000	Hannan-Quinn criter.	-5.955844
F-statistic	1252.350	Durbin-Watson stat	1.456605
Prob(F-statistic)	0.000000		

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 18:07
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.410317	1.700843	-4.356848	0.0004
LNX2	1.628494	0.132485	12.29188	0.0000
LNX5	0.040254	0.025558	1.575049	0.1337

R-squared	0.972226	Mean dependent var	11.00237
Adjusted R-squared	0.968959	S.D. dependent var	0.130543
S.E. of regression	0.023000	Akaike info criterion	-4.569191
Sum squared resid	0.008993	Schwarz criterion	-4.419831
Log likelihood	48.69191	Hannan-Quinn criter.	-4.540034
F-statistic	297.5450	Durbin-Watson stat	0.641682
Prob(F-statistic)	0.000000		

Figure 10. Test results of four binary regression equations

(3) Using lnY=f(lnX2, lnX4) as the base regression equation, three ternary regression equations are established by introducing lnX1, lnX3, and lnX5 respectively, and the results are shown below:

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 18:27
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.537149	1.671614	-5.705354	0.0000
LNX2	0.776481	0.084559	9.182704	0.0000
LNX4	0.938745	0.229742	4.086090	0.0009
LNX1	0.089991	0.053218	1.690982	0.1102

R-squared	0.994281	Mean dependent var	11.00237
Adjusted R-squared	0.993208	S.D. dependent var	0.130543
S.E. of regression	0.010758	Akaike info criterion	-6.049424
Sum squared resid	0.001852	Schwarz criterion	-5.850278
Log likelihood	64.49424	Hannan-Quinn criter.	-6.010549
F-statistic	927.1724	Durbin-Watson stat	1.348931
Prob(F-statistic)	0.000000		

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 18:27
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-10.72093	1.899186	-5.645012	0.0000
LNX2	0.891589	0.184989	4.819674	0.0002
LNX4	0.970620	0.394285	2.461723	0.0256
LNX3	0.066394	0.090267	0.735531	0.4727

R-squared	0.993479	Mean dependent var	11.00237
Adjusted R-squared	0.992256	S.D. dependent var	0.130543
S.E. of regression	0.011488	Akaike info criterion	-5.918254
Sum squared resid	0.002111	Schwarz criterion	-5.719107
Log likelihood	63.18254	Hannan-Quinn criter.	-5.879378
F-statistic	812.5381	Durbin-Watson stat	1.334509
Prob(F-statistic)	0.000000		

ISSN: 2688-9323

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: Least Squares
Date: 12/25/24 Time: 18:28
Sample: 2004 2023
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11.48266	0.985309	-11.65386	0.0000
LNX2	0.620143	0.148612	4.172907	0.0007
LNX4	1.359223	0.180717	7.521293	0.0000
LNX5	-0.018600	0.014637	-1.270754	0.2220

R-squared	0.993877	Mean dependent var	11.00237
Adjusted R-squared	0.992728	S.D. dependent var	0.130543
S.E. of regression	0.011132	Akaike info criterion	-5.981152
Sum squared resid	0.001983	Schwarz criterion	-5.782005
Log likelihood	63.81152	Hannan-Quinn criter.	-5.942276
F-statistic	865.6328	Durbin-Watson stat	1.405357
Prob(F-statistic)	0.000000		

Figure 11. Test results of three ternary regression equations

The t-tests for all three regression equations above failed, so the multiple regression model was developed:

$$LNY=-11.39075+0.772819\ln X2+1.236453\ln X4$$

4.3. Tests for Heteroscedasticity

The following information is obtained using White's test:

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: White

F-statistic	1.979018	Prob. F(2,17)	0.1688
Obs*R-squared	3.777105	Prob. Chi-Square(2)	0.1513
Scaled explained SS	2.189539	Prob. Chi-Square(2)	0.3346

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/25/24 Time: 21:15
Sample: 2004 2023
Included observations: 20
Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.008809	0.005550	-1.587187	0.1309
LNX2^2	-0.000262	0.000133	-1.964707	0.0660
LNX2*LNX4	0.000318	0.000166	1.913987	0.0726

R-squared	0.188855	Mean dependent var	0.000109
Adjusted R-squared	0.093426	S.D. dependent var	0.000142
S.E. of regression	0.000135	Akaike info criterion	-14.84427
Sum squared resid	3.10E-07	Schwarz criterion	-14.69491
Log likelihood	151.4427	Hannan-Quinn criter.	-14.81511
F-statistic	1.979018	Durbin-Watson stat	1.749837
Prob(F-statistic)	0.168786		

Figure 12. The results of the White test

As it can be seen from the above figure, the White's test of the auxiliary regression model its nR2=3.777105 , due to the nR2=3.777105<χ 20.05(2)=5.9915, its p-value is equal to 0.1513 is greater than the level of significance of 0.05, therefore, it does not reject the original hypothesis, and it is considered that the model does not have heteroskedasticity.

ISSN: 2688-9323

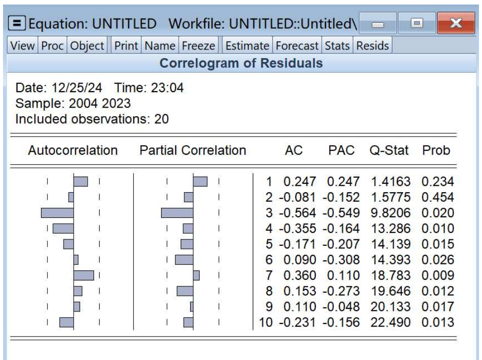
4.4. Autocorrelation Test and Correction

4.4.1. DW Test

Since n=20, k=2, and taking the significance level α=0.05, checking the table of critical values of DW test gives dL=1.100, dU=1.537, and dL< DW=1.45661< dU, it is not possible to determine the existence of autocorrelation by DW test.

4.4.2. Partial Correlation Coefficient Test

By clicking on View/Residual Diagnostics/ Correlogram-Q-Statistics in the Equations window and entering a lag of 10, the period-by-period correlation coefficients and partial correlation coefficients of the residuals vs. are obtained as shown.



	AC	PAC	Q-Stat	Prob
1	0.247	0.247	1.4163	0.234
2	-0.081	-0.152	1.5775	0.454
3	-0.564	-0.549	9.8206	0.020
4	-0.355	-0.164	13.286	0.010
5	-0.171	-0.207	14.139	0.015
6	0.090	-0.308	14.393	0.026
7	0.360	0.110	18.783	0.009
8	0.153	-0.273	19.646	0.012
9	0.110	-0.048	20.133	0.017
10	-0.231	-0.156	22.490	0.013

Figure 13. Residual correlation graph

From the figure, it can be seen that the length of the bars of the third period partial correlation coefficient of this linear model exceeds the dashed portion of the line, so the test results indicate the existence of third-order autocorrelation.

4.4.3. BG Inspection

Click on View/Residual Diagnostics/Serial Correlation LM Test in the Equations window and enter a lag length of 3. The results are shown below.

Equation: UNTITLED Workfile: UNTITLED::Untitled\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3.113486	Prob. F(3,14)	0.0604
Obs*R-squared	8.003663	Prob. Chi-Square(3)	0.0459

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 12/25/24 Time: 23:11
Sample: 2004 2023
Included observations: 20
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.283104	0.840398	-0.336870	0.7412
LNX2	-0.021849	0.078344	-0.278884	0.7844
LNX4	0.045080	0.138694	0.325030	0.7500
RESID(-1)	0.213464	0.223186	0.956442	0.3551
RESID(-2)	0.004308	0.229763	0.018751	0.9853
RESID(-3)	-0.616189	0.227946	-2.703219	0.0171

R-squared	0.400183	Mean dependent var	3.20E-15
Adjusted R-squared	0.185963	S.D. dependent var	0.010718
S.E. of regression	0.009671	Akaike info criterion	-6.196131
Sum squared resid	0.001309	Schwarz criterion	-5.897411
Log likelihood	67.96131	Hannan-Quinn criter.	-6.137818
F-statistic	1.868092	Durbin-Watson stat	2.166138
Prob(F-statistic)	0.164058		

Figure 14. BG test results

ISSN: 2688-9323

The Obs*R-squared in the figure for nR^2 can be seen to correspond to a concomitant probability of 0.0459, which is less than the commonly used significance level of 0.05. Therefore, the model can be considered to be autocorrelated. Both lag one residual (RESID(-1)) and lag two residual (RESID(-2)) failed the significance test, while lag three residual (RESID(-1)) passed the significance test. So by BG test it can be concluded that the model has only third order autocorrelation.

4.4.4. Autocorrelation Correction

The correction for autocorrelation is performed using the generalized difference method by adding an AR term after the OLS estimation command. Since there is a third order autocorrelation, the command is entered as LS LNY C lnX2 lnX4 AR(3). Then in the Estimate Equations window of the above command, click on the Estimate/Option button and select ARMA/Method/GLS (Generalized Least Squares) in the pop-up dialog box, and the results are shown below.

Equation: UNTITLED Workfile: UNTITLED::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LNY
Method: ARMA Generalized Least Squares (BFGS)
Date: 12/25/24 Time: 23:39
Sample: 2004 2023
Included observations: 20
Convergence achieved after 6 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-12.50644	0.579642	-21.57617	0.0000
LNX2	0.736509	0.053288	13.82138	0.0000
LNX4	1.320394	0.095246	13.86305	0.0000
AR(3)	-0.687146	0.200190	-3.432472	0.0034

R-squared	0.995917	Mean dependent var	11.00237
Adjusted R-squared	0.995152	S.D. dependent var	0.130543
S.E. of regression	0.009090	Akaike info criterion	-6.290626
Sum squared resid	0.001322	Schwarz criterion	-6.091479
Log likelihood	66.90626	Hannan-Quinn criter.	-6.251750
F-statistic	1300.928	Durbin-Watson stat	1.812627
Prob(F-statistic)	0.000000		

Inverted AR Roots	.44+.76i	.44-.76i	-.88

Figure 15. The model results were modified after autocorrelation

Due to the existence of third-order autocorrelation, the DW test cannot be used, and the partial correlation coefficient test is used to verify whether the above model eliminates autocorrelation, and the result is that the effect of autocorrelation is eliminated, as shown below:

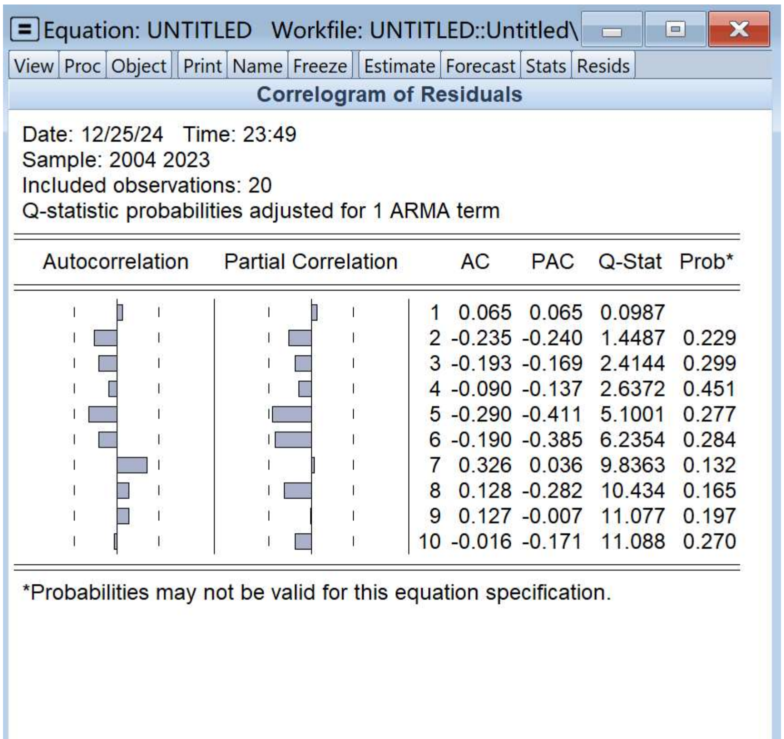


Figure 16. The residual correlation graph after autocorrelation is eliminated

So, the above regression structure is expressed in canonical form as follows:

$$\ln Y=-12.50644+0.736509\ln X2+1.320394\ln X4+\ [AR(3)=-0.687146]$$
$$(13.82138)\ (\ 13.86305)$$
$$R^2=0.9959\ F=1300.928\ DW=1.8126$$

5. Shortcomings of the Model And Directions for Improvement

5.1. Inadequate Modeling

5.1.1. Narrow Selection of Variables

The current model covers only a few factors, such as total power of agricultural machinery, effective irrigated area, the net amount of agricultural fertilizer applied, the sown area of food crops, and the affected area, but does not take into account other factors that may have a key impact on food production, such as changes in temperature and precipitation in the context of climate change, changes in subsidy and purchase policies in the context of agricultural policy adjustments, changes in agricultural science and technology research and development inputs in relation to the cultivation of new varieties and the promotion of new planting techniques, and changes in labor quality. and new planting technologies, and changes in labor quality. This limitation in the selection of variables is likely to make the explanatory power of the model insufficient and incomplete, and unable to accurately reflect the real relationship between food production and the influencing factors.

5.1.2. Inadequate Treatment of Multicollinearity

Although the stepwise regression method is applied to correct for multicollinearity, the method has certain defects and is difficult to completely eliminate the influence of multicollinearity. Moreover, in the process of variable selection, some valuable information will inevitably be lost, which in turn makes the model biased in capturing the intrinsic real relationship between

ISSN: 2688-9323

variables, and fails to accurately present the actual extent of the role of each variable on grain yield.

5.1.3. Difficulties with Autocorrelation Corrections

In correcting for autocorrelation using the generalized difference method, although third-order autocorrelation is eliminated from the results, this does not mean that the model is free of other forms of autocorrelation or serial correlation. In fact, it is likely that there are still hidden autocorrelation problems in the model that have not been detected and dealt with. In addition, the application of the generalized difference method will have an impact on the degrees of freedom of the model to a certain extent, which will bring a potential threat to the stability and accuracy of the model and make the reliability of the model questionable.

5.2. Directions for Improvement

5.2.1. Broadening the Dimensions of Variable Selection

In order for the model to reflect the complex relationship between grain yield and various influencing factors more accurately and comprehensively, the selection of variables should be further expanded to include more factors that may have an impact on grain yield in the model, and more comprehensive data related to them should be widely collected. In this way, it is expected to significantly improve the explanatory ability and forecasting accuracy of the model, so that it can be better adapted to the actual situation and provide more powerful support for the research and forecasting of grain production.

5.2.2. Selection of a more Efficient Multicollinearity Treatment Methodology

Consider adopting more robust multicollinearity treatment methods, such as principal component analysis (PCA) or ridge regression. These methods have the unique advantage of effectively dealing with the problem of multiple covariance on the basis of retaining the original information to the greatest extent possible, so as to more accurately assess the extent of the contribution of each variable in the process of grain yield formation, avoid model bias due to inappropriate selection of variables or loss of information, and further enhance the accuracy and reliability of the model.

5.2.3. In-depth Examination and Treatment of Autocorrelation

More advanced time series analysis methods, such as ARCH (Autoregressive Conditional Heteroskedasticity) model or GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, are used to carry out more in-depth and comprehensive autocorrelation tests and treatments for the model. Through these methods, the dynamic features in the time series data can be mined more carefully to ensure that the model fully takes into account various possible autocorrelations or serial correlations, thus further improving the reliability and stability of the model and enabling it to more accurately characterize and forecast the trend of changes in grain yield.

6. Conclusion and Recommendations

6.1. Conclusion

It was found that effective irrigated area (lnX2) and area sown to grain crops (lnX4) exhibited a significant positive effect on grain yield (lnY) with corresponding coefficients of 0.736509 and 1.320394, respectively.This implies that an increase in the effective irrigated area and the area sown to grain crops, all other things being held constant, will give a strong boost to grain yield .

After a series of rigorous multiple covariance tests and corrections, heteroscedasticity tests and autocorrelation tests and corrections, the constructed model shows high goodness of fit (R2=0.9959) and has good stability. This fully demonstrates that the model can explain the

ISSN: 2688-9323

intrinsic relationship between factors and grain yield more effectively, providing a more reliable tool for the research and analysis of grain yield.

6.2. Recommendations

6.2.1. Strengthening Farmland Water Infrastructure Development

Government departments should further increase investment in the construction of farmland water conservancy facilities, and actively construct and maintain various irrigation channels, reservoirs and other water conservancy projects, so as to effectively increase the effective irrigated area of farmland. Especially in arid areas and water shortage season, stable irrigation conditions play a crucial role in guaranteeing food production. Only by ensuring that crops can obtain sufficient water supply during the growth process can we lay a solid foundation for stable and high grain production.

6.2.2. Scientific Planning and Strict Protection of Arable Land Resources

On the one hand, it is necessary to strictly control the occupation of arable land by construction sites, and to effectively strengthen the protection of arable land, so as to ensure the stability of the area sown by food crops. On the other hand, by actively carrying out land remediation and reclamation, we can increase the area of arable land, optimize the land-use structure, and improve the efficiency of land use, so as to provide sufficient land resources for food production, realize the sustainable use of arable land resources, and maintain the stable development of food production.

6.2.3. Greatly Promote Agricultural Science and Technology Innovation and Application

Agricultural research institutes and enterprises are actively encouraged to increase their investment in research and development in agricultural science and technology, and are committed to breeding high-quality food varieties that are high-yielding, resistant to diseases and pests, and adapted to different environmental conditions. At the same time, advanced planting techniques and management experience, such as precision farming techniques and intelligent irrigation systems, are being widely promoted. Through these initiatives, it is possible not only to significantly improve agricultural production efficiency and the level of grain yields, reduce excessive reliance on traditional inputs of factors of production, and lower production costs, but also to realize the sustainable development of grain production and enhance the competitiveness and risk-resistant capacity of China's grain industry.

6.2.4. Building a Sound Early Warning and Response Mechanism for Agricultural Disasters

Strengthening meteorological disaster monitoring and early-warning capacity building, continuously improving the accuracy and timeliness of disaster monitoring, and ensuring that disaster information can be released in a timely and accurate manner, so as to provide strong support for farmers to take precautionary measures in advance. In addition, the government should further increase support for agricultural insurance, continue to improve the agricultural insurance system, reduce the risk of farmers damaged by disasters, and fully guarantee the production enthusiasm of farmers and income stability. Through the construction of a comprehensive agricultural disaster early warning and response mechanism, to ensure that food production in the face of natural disasters and other adverse factors, but still be able to maintain relative stability, to minimize the impact of disasters on food production.

References

[1] Kong Fanjin. Measures and effects of developing modern agriculture in China under the background of rural revitalization [J]. Agricultural Economy 2023, (08):28-29.

ISSN: 2688-9323

[2] Zhang Zhixin, Gao Mingtao, Zhang Xiujie. High standard farmland construction, new quality agricultural productivity and grain increase [J]. Southwest Finance,2024,(10):17-29.

[3] Wang Lingling. Applying chemical fertilizer and pesticide content of research on the effects of grain production in China [D]. Dalian ocean university, 2024. The DOI: 10.27821 /, dc nki. Gdlhy. 2024.000366.

[4] Yang Xianliang. Research on Coupling coordination of agricultural mechanization and food security [D]. Northwest A&F University,2024.

[5] Zhang Chenxi, Li Huiling, Chen Gaoko & Aubrey Talip. Research on dynamic evolution of grain production efficiency in China: An empirical analysis since the reform and opening up. Grain, oil and food Science and technology 1-10.

[6] Yu Ao, Li Jin, Jia Zhuoqiang. Research on mechanism and path of intelligent agricultural mechanization to ensure food security [J]. Rural Economy,2024,(11):33-44.