Analysis of China's Grain Economy Based on CEIC Economic Database

Mengyao Li
Hunan University of Finance and Economics, Changsha 410000, China.

Abstract: In recent years, 70% of China's population is rural population, and the development of agricultural production is directly related to the improvement of farmers' living standards and the realization of national economic construction goals. There are many factors affecting grain production, and this paper will analyze some of the factors affecting grain production in China, and use SPSS statistical software and stepwise regression analysis method to establish a regression model of grain production in China, from which the main influencing factors are separated.

Keywords: Economic Development; Economic Database; Grain Production

1. Introduction

The study showed that the model established by using stepwise regression analysis had a good fitting effect, and the main factors affecting grain yield in China were: fertilizer application and grain crop cultivation area. The study concluded that increasing the cultivated area of grain crops is the most effective way to increase grain yield.

2. Data presentation and description

Analysis: There are many factors affecting total grain production, including the area under cultivation of grain crops, grain area per unit yield, effective irrigation area, fertilizer usage, pesticide usage, total power of agricultural machinery, etc. Three explanatory variables are selected: agricultural fertilizer application (X2), total grain sown area (X3) and total rural population (X4) (due to agricultural labour The total grain production (Y) in China from 1990 to 2016 was analysed and the model developed was quantified using econometric methods to study the degree of influence of each influencing factor.

Table 2: Total grain output, total fertilizer application, total grain sown area, total rural population				
The given year	Total grain output (Y)/ thousand tons	Total fertilizer application (X2)/ thousand tons	Total sown area of grain (X3) per thousand hectares	Total rural population (X4)/ million
1983	387275	16598	114047.3	807.34
1984	407305	17398	112884	803.4
1985	379108	17758	108845.1	807.57
1986	391512	19306	110932.7	811.41
1987	402980	19993	111268	816.26
1988	394080	21415	110122.7	823.65
1989	407550	23571	112204.7	831.64
1990	446243	25903	113465.9	841.38
1991	435293	28051	112313.6	846.2
1992	442658	29302	110559.7	849.96
1993	456488	31519	110508.7	853.44
1994	445101	33179	109543.7	856.81
1995	466618	35937	110060.4	859.47
1996	504535	38279	112547.9	850.85
1997	494171	39807	112912.1	841.77
1998	512295.3	40837	113787.4	831.53
1999	508385.8	41243.2	113161	820.38
2000	462175.2	41464.12	108462.5	808.37
2001	452636.7	42537.63	106080	795.63
2002	457057.5	43393.9	103890.8	782.41
2003	430695.3	44115.6	99410.37	768.51
2004	469469.5	46365.8	101606	757.05
2005	484021.9	47662.18	104278.4	745.44
2006	498042.3	49276.93	104957.7	731.6
2007	501602.8	51078.32	105638.4	714.96
2008	528709.2	52390.23	106792.7	703.99
2009	530820.8	54043.52	108985.8	689.38
2010	546477.1	55616.8	109876.1	671.13
2011	571208.5	57042.36	110573	656.56
2012	589579.7	58388.49	111204.6	642.22
2013	601938.4	59118.64	111955.6	629.61
2014	607026.1	59959.38	112722.6	618.66
2015	621439.2	60226.03	113342.9	603.46
2016	616250.5	59840.29	113034.5	589.73

Data source: CEIE Economic Database

3. Descriptive analysis

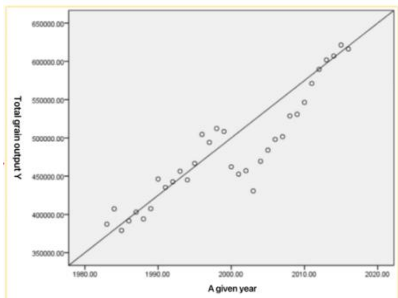
Firstly, based on a comparison of national and provincial (district and municipal) grain production in 2016.

		Sown area/ thousand hectares	Yield per unit area (ha) /thousand ha
Sown area/thousand hectares	Pearson correlation significance (two-tailed)	1	.075 .690
	N	31	31
Yield per unit area (ha)/thousand ha	Pearson correlation significance (two-tailed)	.075 .690	1
	N	31	31

Analysis:

Bivariate correlation correlation analysis (correlation between sown area and yield per unit area and total yield), we can get the correlation between sown area and yield per unit area by bivariate correlation analysis two-tailed significance test, significance 0.690 is less than 1.000, indicating that there is a correlation between these two scale variables and that they are strongly correlated, however, by this result we cannot conclude that There is a causal relationship between the two. Again, using cluster analysis (a hierarchical clustering analysis of yields per unit area for 31 different regions and provinces after normalising the data):

The dependent variable in this case is total grain production and its distribution is plotted on a scatter plot



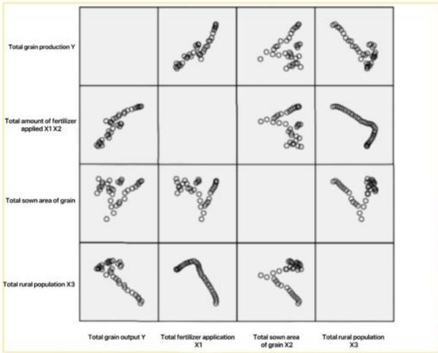
Scatter plot of total grain production in different years

As can be seen from the graph, China's total grain production in different years is on an upward trend according to the arrangement of each year.

4. Modelling analysis

4.1 Model setting

First, the parameters of the model were analysed and estimated using SPSS software based on relevant data from 1983-2016 to obtain a matrix plot of the series Y, X1, X2 and X3.



It can be seen that the total food production and the various influencing factors vary significantly, and their changes are basically in the same direction, and may have some correlation with each other, setting the model in the form of a linear regression model:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\mu$$

4.2 Estimated parameters

A linear regression analysis was performed on the above data using SPSS to estimate the model parameters and the outputs 1, 2 and 3 are shown below.

Figure 1

coefficient

model	Nonnormalized coefficient		Standardization coefficient	T	significance	95.0% confidence interval of B	
	B		Beta			lower limit	Upper limit
1 (constant)	-422725.086	68261.143		-6.193	.000	-562132.939	-283317.234
Total fertilizer application/thousand tons	4.611	.208	.936	22.147	.000	4.186	5.036
Total area/thousand hectares	7.023	.467	.368	15.036	.000	6.069	7.977
Total rural population/million	-64.024	34.662	-.076	-1.847	.075	-134.813	6.765

a. Variable: total grain production/thousand tons

Figure 2

Model summary

Model	R	R squared	Adjusted R squared	Standard skew error	Durbin-Watson
1	.992[a]	.984	.982	9405.25912	1.553

a. Predicted value: (constant), total rural population/million people, total area/thousand hectares, total fertilizer application/thousand tons

b. Variable: total grain production/thousand tons

Analysis: First, the model is tested as a whole by looking at the p-value corresponding to the F-value. The p-value is less than 5%, i.e. the model is significant and the model is meaningful. Then, look at the p-value corresponding to the t-value. The p-value is less than 5%, i.e. the coefficient is significant and the coefficient is not zero.

Figure 3

Residual statistics

	Minimum value	Maximum value	Mean number	Standard deviation	N
Predicted value	371849.7500	612328.4375	483845.5453	69772.51576	34
Residual error	-16899.14648	14828.27051	.00000	8967.56271	34
Standard predicted value	-1.605	1.841	.000	1.000	34
Standard residual	-1.797	1.577	.000	.953	34

a. Variable: total grain production/thousand tons

(1) Based on the output it can be concluded that the model estimation is written as

$$Y=-422725.086+4.611X_1+7.023X_2-64.024\ X_3$$

（68261.143）（0.208）（0.467）（34.662）

T=（-6.193）（22.147）（15.036）（-1.847）

R-squared = 0.984, with a modified decidability coefficient of 0.982, which indicates that the model fits the sample well.

F=605.369 DW=1.533

(2) R=0.992 and the coefficient of determination R-squared=0.984, which shows that the regression equation is highly significant by the coefficient of determination.

(3) From the ANOVA table it can be concluded that F=605.369, p-value=0.000, indicating that the regression equation is highly significant, indicating that X1, X2 and X3 as a whole have a highly significant linear effect on Y.

5. Model testing

5.1 Economic significance test

In terms of economic significance, China's total grain production Y is positively correlated with total fertilizer application X1, total grain sown area X2, and negatively correlated with total rural population X3.

5.2 Statistical tests

(1) Goodness-of-fit test. The regression results show that the values of the square of R and the adjusted square R are both close to 1, indicating that the model has a good fit.

(2) t-test. It can be seen from the table: at the significance level of $\alpha=0.05$, the critical value of the t-statistic for the degree of freedom n-k-1=30 is $t\alpha/2(30)=3.03$, the t-values of X1 and X2 are greater than this critical value, and the t-value of X3 is less than this critical value, so X1 and X2 are significant at the 95% level and pass the variable significance test, while X3 is not significant at the 95% level and does not X3 was not significant at the 95% level and did not pass the variable significance test.

(3) F test. The critical value of F statistic is F0.05 (3, 30) = 2.922, F is greater than this critical value, so the linear relationship of the model is significant at 95% confidence level.

5.3 Testing of the regression model

5.3.1 Test for multicollinearity

Figure 4

coefficient

Model	Nonnormalized coefficient		Standardization coefficient	T	significance	Collinear statistics	
	B	Standard error	Beta			tolerance	VIF
1 (constant)	-422725.086	68261.143		-6.193	.000		
Total fertilizer application/thousand tons	4.611	.208	.936	22.147	.000	.303	3.297
Total area/thousand hectares	7.023	.467	.368	15.036	.000	.904	1.106
Total rural population/million	-64.024	34.662	-.076	-1.847	.075	.317	3.152

As can be seen from the output 4, the variance expansion factors VIF for X1, X2 and X3 are 3.297, 1.106 and 3.152 respectively, all three numbers are less than 10 and the regression coefficients all have a reasonable economic interpretation, indicating that there is no multicollinearity in this regression model and it can be used as the final regression model. The regression equation is

Y=-422725.086+4.611X1+7.023X2-64.024 X3

5.3.2 Heteroskedasticity test

An ordinary least squares regression of Y on X1, X2 and X3 was created using SPSS software and the residuals were retained and the results are shown in outputs 5 and 6 below.

Figure 5

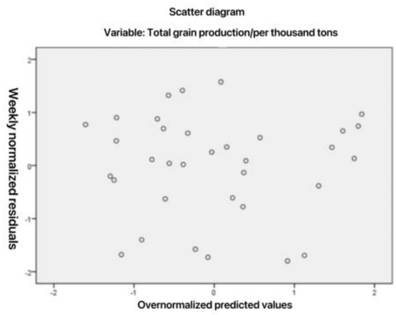
Variance analysis

model		Sum of squares	df	Mean square	F	significance
1	regression	1.607E+11	3	5.355E+10	605.369	.000b
	Residual error	2653766972	30	88458899.06		
	total	1.633E+11	33			

a. Variable: total grain production/thousand tons

b. Predicted value: Constant, total rural population/million people, total area/thousand hectares, total fertilizer application/thousand tons

Figure 6



It is generally accepted that if a regression model satisfies the basic assumptions given, so the residuals should vary randomly around e=0 and within a region where the variation is not significant, so the residual plot shows that there is no heteroskedasticity in the model.

5.3.3 Autocorrelation test.

Using DW test, from Figure 2 can be seen that the regression equation coefficient of determination X1, X2 significant, for n = 34, k = 3, a = 0.05, DW test statistic = 1.553, check DW statistics table can be seen, dl = 1.27, du = 1.65, so du = 1.27 < DW < 4 - du = 2.73, indicating that the model does not exist autocorrelation.

5.3.4 Determination of the model

Through the above tests, the model was finally determined as:

$$Y=-422725.086+4.611X_1+7.023X_2-64.024 X_3 +u$$

6. Conclusion

The results of the above analysis show that of the three factors selected, the effect of total fertilizer application and total area sown to grain was more significant. From the regression model it can be seen that the most significant contribution to grain yield is made by the amount of fertilizer applied. Although the contribution of sown area to grain production is not as significant as that of fertilizer application, the increase in arable land area contributes more to the increase in grain production. The coefficient of influence of total village population is smaller, but if the absolute value of total village population is larger, then it will cause a larger decrease in grain production, so balancing the number of village population is the key to increasing grain production.

References

[1] Nancy L. Leech Karen C. Barrett Ceorge A. Morgan SPSS for Intermediate Statistics Use and Interpretation, Third Edition, Publishing House Of Electronics Industry; 2009.

[2] Vasicek, O. A., & Fong, H. G. (1982). Term Structure Modeling Using Exponential Splines. The Journal of Financ e, 37(2), 339–348.

[3] Ait-Sahalia, Y. (1996). Testing Continuous-Time Models of the Spot Interest Rate. The Review of Financial Studies, 9(2), 385–426.