

A Regression Analysis of Chinese Agriculture in Shanxi Province

Juzhen Wu

Department of Mathematics, Jiangnan University, Wuhan 430056, China

juzhen_wu@163.com

Keywords: Agriculture; Regression analysis; Shanxi province; County-level data

Abstract. This study is aimed to investigate which factors can impact regional agricultural output. We use 120 county-level data to do regression analysis of Chinese agriculture in Shanxi province. The result shows that rainfall, tractors, fertilizers usage, proportion of irrigated land all have significant impact on total agricultural output. In addition, this production function is not constant returns to scale. The econometric results also suggest that in order to increase the agricultural output in Shanxi province, the most efficient way is to increase the fertilizer usage and irrigated farmland. This study is very helpful for public policy to improve agricultural efficiency in Shanxi providence of China.

Introduction

China is the world's largest producer and consumer of agricultural products. In the past, Chinese governmental policy advocated regional self-sufficiency in agricultural production [1-4]. Cho et al. studied the output growth in Chinese agriculture and they reviewed the evidence regarding the growth [5]. Using provincial-level production data, Fan found that investment in agricultural research accounted for 20 percent of productivity growth [6]. Fan and Pardey found that 38% of the production growth was due to in policy impact and 8% due to good weather condition during 1980-1985 [7]. There are many other literatures focusing on Chinese agricultural efficiency and growth. However, the data that most of the papers used are provincial-level aggregated data.

This study is aimed to investigate which factors can impact regional agricultural output. I use 120 county-level data to do regression analysis of Chinese agriculture in Shanxi province. We know that the agricultural production can be effect by many factors, such as land, labor, capital, electricity, rainfall, and average income. This regression analysis is helpful to identify which factors are able to enhance the efficient usage of agricultural resources. It is also very helpful for public policy to improve agricultural efficiency in Shanxi providence of China.

This study extends the literature in two important ways. First, it carries out a regression analysis using county-level data. Second, it firstly introduces six kinds of agricultural inputs as independent variables in the agricultural study of Shanxi providence. The rest of the paper unfolds as the following. Section 2 introduces the methodologies we use in the study, followed by the results and empirical analysis in section 3. Section 4 concludes.

Data

The data of this study comes from Department of Agriculture and Shanxi Meteorological Bureau. This data set contains detailed information of 120 counties on total agricultural output (in dollar value), agricultural labor (number), total farm land (hectare), irrigated land (hectare), total mechanical power (kilowatt), fertilizer usage (ton), agricultural electricity usage (10 thousand kilowatt), and annual average rainfall (millimeter). Among those 120 counties, only 106 counties have valuable and complete data including all variables. Therefore, in this study, I conduct the analysis by looking at these 106 counties.

We are able to construct the quantities of inputs and output for each county. We make the dependent variable as the total agricultural output (Y) and the independent variables as the agricultural inputs. There are five independent variables. They are labor(X1), land(X2), mechanical power(X3), fertilizer(X4), and electricity(X5). In addition, some other explanatory variables I also considered in this regression analysis are annual rainfall and the proportion of irrigated land.

Methodology

An important consideration in studies using the regression analysis is to define a clear approach to be followed when searching the relevant literature. A Cobb–Douglas production function was estimated for agricultural production in China [8]. The Cobb–Douglas function form can be estimated as a linear relationship using the following expression:

$$\ln(Y) = \beta_0 + \sum_i \beta_i \ln(I_i) + u \tag{1}$$

Where: Y is the total agricultural output
 I_i are all kinds of inputs variables
 β_i are the coefficients

In this study, I also considered two explanatory variables in regression analysis. They are annual rainfall and the proportion of irrigated land. Adequate rainfall is an essential requirement for agricultural harvest. So it is considerable variable across different counties of the providence. Irrigated land would enhance input efficiency and make the agricultural production more stable. Hence, proportion of irrigated land also impact on the total agricultural output. Therefore, the adjusted overall linear relationship equation is shown in the following:

$$\ln(Y) = \beta_0 + \sum_i \beta_i \ln(I_i) + \sum_j \alpha_j x_j + u \tag{2}$$

Where: Y is the total agricultural output
 I_i are all kinds of inputs
 β_i, α_j are the coefficients
 X_j are explanatory variables

In the analysis, I will conduct several alternative runs containing different variables and compare the results between those runs. Then, I will explain and discuss those results. Moreover, for the production function, I will do the test of constant return to scale and discuss the results. The term returns to scale arises in the context of a firm's production function. It explains the behavior of the rate of increase in output (production) relative to the associated increase in the inputs (the factors of production) in the long run [9]. In the long run all factors of production are variable and subject to change due to a scale.

Results and Analysis

At first, we only consider five input variables. Using equation (1) with 5 inputs variables (labor, total land, mechanical power, fertilizer, and electricity), we can see that both fertilizer usage and electricity usage have significant impact on the agricultural output. This means increase the use of fertilizers and electricity both can increase the output. However, in this model, the land has negative impact on output. Ever though it is not significant from the t-test, the negative sign of coefficient of labor does not make sense. So we need to an alternative run and compare the results.

Fig. 1, shows that the results of a test of constant returns to scale. Our hypothesis of the test is the sum of all coefficients equals to one. From the P value, we know that the hypothesis is rejected and the production function does not have constant returns to scale.

```

loglabor + logland + logpower + logfertilizer + logelectricity = 1

F( 1, 100) = 0.09
Prob > F = 0.7618

```

Figure 1. Test of Constant Returns to Scale (1)

Now, we separate the land variable into two variables. They are irrigated land and non-irrigated land. For equation (2) with 6 inputs variables (labor, irrigated land, non-irrigated land, mechanical power, fertilizer, and electricity) and two explanatory variables (rainfall and proportion of irrigated land), using STATA, we can see that fertilizer usage, irrigated land, and proportion of irrigated land has significant impact on the agricultural output. However, in this model, the sign of irrigated land factor is negative, which means an increase in irrigated land would decrease the output. This makes no sense at all. When we look at the variables, we find that irrigated land and proportion of irrigated land are correlated with each other. This means we have an endogenous problem here [10].

This time, we use total land as one variable in another alternative run. For equation (2) with 5 inputs variables (labor, total land, mechanical power, fertilizer, and electricity) and two explanatory variables (rainfall and proportion of irrigated land), use STATA, shows that all variables have positive signs on coefficients. R² equals 0.7829. Land, rain, and the proportion of irrigated land are all significant. The P-value of electricity is higher than the two runs before. In agriculture, since electricity is mainly used to pump water and irrigate farmland. The more amount of rainfall, the less need for electricity. So, rainfall and electricity are very likely to have correlation. In the next run, I will drop electricity and compare the results.

```

loglabor + logland + logpower + logfertilizer + logelectricity = 1

F( 1, 98) = 0.43
Prob > F = 0.5159

```

Figure 2. Tests of Constant Returns to Scale (2)

Our hypothesis of testing Constant Returns to Scale is the sum of all coefficients equals to one. From the P value, we see that the hypothesis is rejected and the production function does not have constant returns to scale.

Finally, we use total land as one variable and we also drop the electricity. Using equation (2) with 4 inputs variables (labor, total land, mechanical power, and fertilizer) and two explanatory variables (rainfall and proportion of irrigated land), shows us that all variables have positive signs on coefficients. R² equals 0.7819. Root MSE in this run is less than the previous run. Now, mechanical powers, fertilizers, rain, the proportion of irrigated land are all significant. Coefficient of proportion of irrigated land is almost equal to 1. Logfertilizer has the second highest value of coefficient.

```

loglabor + logland + logpower + logfertilizer = 1

F( 1, 99) = 0.53
Prob > F = 0.4664

```

Figure 3. Tests of Constant Returns to Scale (3)

Our hypothesis of testing Constant Returns to Scale is the sum of all coefficients in this run equals to one. This time, we have a lower value of P value, but we still reject the hypothesis that the production function does not have constant returns to scale.

Before conducting the regression analysis, we were expecting that labor, land, capital, rainfall and proportion of irrigated land have positive effect on agricultural output. Now, it comes out that even though they all have positive effect, but labor and land are not significant. Results show that rainfall is significant. However, the value of coefficient is so small that a slight increase of rain could not change output much. Fertilizer usage and irrigated farmland play the most important role for the changes of agricultural output in Shanxi province.

Summary and Conclusions

The objective of this study was to undertake a regression analysis seeking to explain the variation in output focusing on the agricultural sector. We include 106 county-level data to estimate linear relationship follow the Cobb–Douglas function. In addition, alternative models incorporated explanatory variables to capture other effects on agricultural output. This study contributes to the county level or regional level productivity literature because the existing work in this area typically uses aggregate (i.e., national) level data to estimate factors on output and productivity. The results show that rainfall, tractors, fertilizers usage, proportion of irrigated land all have significant impact on total agricultural output. In addition, this production function is not constant returns to scale. The econometric results also suggest that in order to increase the agricultural output in Shanxi province, the most efficient way is to increase the fertilizer usage and irrigated farmland. This study is very helpful for public policy to improve agricultural efficiency in Shanxi providence of China.

For the future study, I would like to see what happens in different years and compare the results. I also would like to use panel data to conduct the analysis and see whether I can find something interesting.

References

- [1] Ren J., Yu Y. and Wang R., Preliminary Analysis of the Environmental Impact of Chinese Agricultural Policy, *Chinese Agricultural Science Bulletin*, 25(15), 223-229. (2009)
- [2] Zong Y., Li X. and Qiao L., An Empirical Study on Agricultural Support of China's Agricultural Policy, *Scientia Agricultura Sinica*, 40(3), 622-627. (2007)
- [3] He S., On the agricultural policy effects in China, *Statistical Research*, 29(1), 43-48. (2012)
- [4] Li Y., The adjustment of Chinese agricultural policies in post-WTO period, *Journal of Nanjing Agricultural University*, 24(1), 99-102. (2001)
- [5] Cho, S., Chen, Z. and Poudyal, N. C., Spatial structure of agricultural production in China, *Applied Economics* (DOI: 10.1080/00036840701765320). (2008)
- [6] Fan, S., Effects of technological change and institutional reform on production growth in Chinese agriculture. *American Journal of Agricultural Economics*, 73 (2), 266-275. (1991)
- [7] Fan S. and Pardey P.G., Research, productivity, and output growth in Chinese agriculture. *Journal of Development Economics*. (1997)
- [8] Wan, G. H. and Cheng, E., Effects of land fragmentation and returns to scale in the Chinese farming sector, *Applied Economics*, 33, 183–94. (2001)

[9] Frisch, R., Theory of Production. Chicago: Rand-McNally. (1965)

[10] Greene W., Econometric analysis, 6th ed. Pearson Education. (2002)