Technical efficiency analysis of major agriculture production provinces in China: a stochastic frontier model with entropy

Xi Ai¹, Jianxu Liu¹,²,³, Jirakom Sirisrisakulchai¹,² and Songsak Sriboonchitta¹,²

¹Faculty of Economic, Chiang Mai University, Chiang Mai 50200, Thailand
²Puey Ungphakorn Center of Excellence in Econometrics, Chiang Mai University, Chiang Mai 50200, Thailand
³Email: liujianxu1984@163.com

Abstract. Agriculture plays an important role in Chinese economy. This study aims to analyze the agricultural technical efficiency of top 10 agricultural provinces in China. Also, a classical Stochastic Frontier Model (SFM) and a Stochastic Frontier Model (SFM) with entropy are applied in to this study. In our empirical results, we find that the Stochastic Frontier Model (SFM) with entropy is much better than the classical SFM in terms of economic interpretation. From the perspective of technical efficiency, Guangdong Province has been approaching the frontier of agricultural technology efficiency in the past six years. From a national perspective, this means that there is still much room for improvement in technical efficiency and productivity in other provinces. Therefore, we suggest that provinces with technical inefficiency should learn from sample provinces to increase agricultural productivity.

1. Introduction

Agriculture is the basic industry and strategic industry of the national economy, and it is directly related to national food security and stable development of the economy. China has completely abolished agricultural taxes since 2006, and has achieved important achievements in lightening the burden on farmers and realizing “industry returns to help the agriculture and city supports the countryside”. In spite of this, the agricultural aspect of some provinces still faces problems such as low technical efficiency, low productivity, high production costs and low output and so on. It is difficult for rural ecological environment to bear the traditionally extensive planting mode with high input and high consumption. Under the new circumstances, whether China’s agricultural production can continue to grow has become the focus of widespread concern in the government and academia. Therefore, it is imperative to increase technical efficiency and productivity. In order to solve the above problems, we choose the stochastic frontier model (SFM) with generalized maximum entropy (GME) to analyze the factors affecting agricultural added value in 10 provinces of Shandong, Henan, Jiangsu, Hebei and Sichuan, and calculate the technical efficiency of the sample provinces. Through the analysis of the developed agricultural provinces, it is hoped that they will provide practical and effective experiences for underdeveloped provinces in agriculture, and to change the current situation of the slow growth of agricultural production in these provinces. The rest of this paper proceeds as follows. Section 2 describes the contribution of the literature. Section 3 explains the SFA and GME models, and analyzes the data and variables, as well as the calculation of technical efficiency. Section 4 analysis the results of GME and MLE. Finally, Section 5 summarizes the full text.
2. Literature review

The measurement of agricultural efficiency has always been of interest to agricultural economists. The stochastic frontier model (SFM) proposed independently by Aigner et al. and Meeusen and Van den Broeck, is also rapidly become the most commonly used model for technical efficiency analysis, and has been widely applied in many technical fields, especially in agricultural economics [1,2]. For example, Tian and Li used the stochastic frontier model to examine technical efficiency and the technology gap in China’s rapeseed production [3]. Liu et al. analyzed the 300 farmers in north central Thailand through the zero inefficiency stochastic frontier model (ZISFM) and looked for the decisive factors affecting their technical inefficiency. The results show that education has the greatest impact on improving efficiency [4]. Ma et al. used the translog stochastic frontier production model to analyze China’s provincial agricultural productivity, it is proposed that high population density is the main reason for technical inefficiency [5].

Because of the shortcomings of SFM and maximum likelihood estimation (MLE), the economists proposed a generalized maximum entropy (GME) boundary estimation combining the advantages of these two methods of analysis. Sen Gupta applied maximum entropy to derive the distribution of unilateral error terms in the production frontier model [6]. Zhang and Fan used the GME method for the first time to empirically estimate China’s agricultural multi-output production functions and input allocations, and proposed correct understanding of grain production technology is essential [7]. Rezek and Campbell used GME to estimate the shadow prices for power plant emissions [8]. Campbell et al. conducted a comparative analysis of GME, SFM, and DEA [9]. Pedro and Elvira explored the technical efficiency of the Portuguese wine region through GME estimates. The results show that the production units are moving away from the production boundary [10].

3. Econometric methodology

In this section, we first summarize and review the stochastic frontier production model and estimate the technical efficiency (Subsection 3.1). Then GME was introduced in Subsection 3.2. Finally we explained the sample and data in Subsection 3.3.

3.1. Stochastic frontier model

This study used the stochastic frontier production model to estimate and analyze the agricultural industry technical efficiency in China’s top 10 agricultural production provinces. Consider the following translog stochastic frontier production model:

\[
\ln y_{it} = \alpha_0 + \sum_{j=1}^{5} \alpha_j \ln x_{j, it} + \frac{1}{2} \sum_{j=1}^{5} \sum_{k=1}^{5} \alpha_{jk} (\ln x_{j, it}, \ln x_{k, it}) + v_{it} - u_{it} \tag{1}
\]

where,

- \(y_{it}\) denotes output, the added value of agriculture (including agriculture, forestry, animal husbandry and fisheries);
- \(x_{1i}\) denotes labor, the number of employees in the agricultural industry;
- \(x_{2i}\) denotes land, the area of cultivated area and non-cultivated area of crops;
- \(x_{3i}\) denotes machinery, the total power of sizes tractors, irrigation and drainage motors, combine harvesters and other machines;
- \(x_{4i}\) denotes fertilizer, the amount of fertilizer used;
- \(x_{5i}\) denotes geomembrane, the plastic membrane for agriculture used;
- \(\alpha_j\) denotes the parameters of the input elasticity on output;
- \(\alpha_{jk}\) denotes the alternative elasticity between input \(j\) and input \(k\), where \(\alpha_{jk} = \alpha_{kj}\);
- \(v_{it}\) is an error term. We assume that the distribution of random error term \(v_{it}\) is \(v_{it} \sim N(0, \sigma_v^2)\);
- \(u_{it}\) is a positive error term. We assume that the positive error term \(u_{it}\) has a non-negative truncated distribution of \(u_{it} \sim N(0, \sigma_u^2)\).

The output-oriented technical efficiency is measured by the equation as follows:
By the virtue of maximum likelihood estimation (MLE), in order to facilitate the estimation, we can refer to Battese and Corra to construct variance parameters [11]. Replacing $\sigma_i^2$ and $\sigma_u^2$ with $\sigma^2 = \sigma_i^2 + \sigma_u^2$ and $\gamma = \sigma_u^2 / (\sigma_i^2 + \sigma_u^2)$, The value of the parameter is $0 < \gamma < 1$. The statistical test of $\gamma$ estimates can reflect whether the variation in the technical efficiency of the production unit is significant. When $\gamma$ approaches 0, it means that the random error dominates, and the significant difference in technical efficiency does not exist at this time. When $\gamma = 0$, it means that the inefficiency comes entirely from the influence of random factors. When $\gamma$ approaches 1, the technical inefficiency dominates. In addition, the equation of technical inefficiency can be written as follows:

$$U_{it} = \delta_0 + z_{it}\delta_m + w_{it}$$

where $U_{it}$ is the technical inefficiency, $\delta_0, \ldots, \delta_m$ are the parameters to be estimated, $z_{it}$ is the determinant factor of technical inefficiency, $w_{it}$ is the random error.

3.2. The generalized maximum entropy (GME) estimator

Consider the following general linear model

$$y = X\beta + \nu$$

where $y$ is the $(N \times 1)$ vector of the observed variable values, $X$ is the known $(N \times K)$ matrix, $\beta$ is the $(K \times 1)$ vector of unknown parameters, and $\nu$ is the random disturbances vector. By modifying the generalized maximum entropy method described by Golan et al. [12]. We use each $\beta_k$ as a discrete random variable, there is a set of $K$ support points ($q_k$) and probability weights ($p_k$) for each observation, with a compact support and $2 \leq J \leq \infty$ possible outcomes. In short, each parameter equals the product of the support point and its associated probability weights and is summed over all support points. The linear model in (4) can be expressed as:

$$y = XQp + G\omega$$

where,

$$\beta = Qp = \begin{bmatrix} q_1' & 0 & \cdots & 0 \\ 0 & q_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q_k' \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix}$$

where $\beta$ is $K \times KJ$ and $p$ is $KJ \times 1$ ($p \gg 0$), there is a set of $T$ support points ($g_1$) and probability weights ($\omega_1$) for each observation, with $2 \leq T < \infty$. The random disturbance vector is consider the following general linear model

$$\nu = G\omega = \begin{bmatrix} g_1' & 0 & \cdots & 0 \\ 0 & g_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_k' \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \end{bmatrix}$$

where $\nu$ is $N \times NT$ and $\omega$ is $NT \times 1$ ($\omega \gg 0$), the unknown vectors of probability weights $p$ and $\omega$ unknown vectors, are estimated by maximizing the entropy function:

$$\max H(p, \omega) = -p' \ln(p) - \omega' \ln(\omega)$$

Subject to the model constraint and the additivity constraints on $p$ and $\omega$,

$$y = XQp + G\omega$$

$$i_K = (I_K \otimes i'_K)p$$

$$i_N = (I_N \otimes i'_N)\omega$$
where ⊗ represents the Kronecker product. I represents the identity matrix, and i is a column. The point estimation used to form unknown parameter vectors (and the unknown random errors) is that the GME estimator generates the optimal probability vector \( \hat{\rho} \) and \( \hat{\omega} \).

In order to define it as the two-sided error component, we have extended the GME method to include a one-sided inefficiency component \( u \). For each \( u_i \), there is a set of \( M \) support points \( (h_i) \) and probability weights \( (\phi_i) \) for each observation, with \( 2 \leq M < \infty \). The lower bound of the support point for the one-sided inefficiency component is zero for all observations, which is the opposite of the two-sided disturbances. All other support points are positive:

\[
\begin{align*}
h_{i1} &= 0 \quad \forall \, i \\
h_{im} &> 0 \quad \forall \, i \text{ and } m \geq 2
\end{align*}
\]

In matrix form

\[
u = H\phi = \begin{bmatrix} h'_1 & 0 & \cdots & 0 \\ 0 & h'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h'_N \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix}
\]

where \( u \) is \( N \times NM \) and \( \phi \) is \( NM \times I \) \((\phi \gg 0)\).

Combining equation 5-14, the general linear model to be estimated is

\[
y = XQp + G\omega - H\phi
\]

Estimate the probability weights \( p \), \( \omega \) and \( \phi \) unknown vectors by maximizing the entropy function:

\[
\max H(p, \omega, \phi) = -p' \ln(p) - \omega' \ln(\omega) - \phi' \ln(\phi)
\]

Subject to the model constraint and the additively constraints on \( p \), \( \phi \) and \( \omega \)

\[
y = XQp + G\omega - H\phi
\]

\[
i_k = (I_K \otimes i'_j)p
\]

\[
i_N = (I_N \otimes i'_R)\omega
\]

\[
i_N = (I_N \otimes i'_M)\phi
\]

where \( \otimes \) represents the Kronecker product, \( I \) represents the identity matrix, and \( i \) is a column. The point estimation used to form unknown parameter vectors (and the unknown random errors) is that the GME estimator generates the optimal probability vector \( \hat{\rho} \), \( \hat{\omega} \) and \( \hat{\phi} \).

3.3. The data

Table 1 summarizes the variables in the stochastic frontier model. All data indicators come from the National Bureau of Statistics of China (NBSC), including China’s top ten agricultural production provinces (Shandong, Henan, Jiangsu, Hebei, Sichuan, etc.) from 2007 to 2012, and provided us with 60 observations. Table 1 shows that the dependent variable is agricultural output. The independent variables are land, labor, mechanical power, fertilizers, and geomembranes. The labor represents the number of agricultural employees. The land (sown area) refers to the cultivated area and non-cultivated area of the crop. The mechanical power refers to the total power of the tractor, irrigation drainage motor and combine harvester. The fertilizer refers to the amount of fertilizer. The geomembrane refers to the number of plastic films used in agriculture. The summary statistics of variables in the stochastic frontier model are given in table 1. The average value of 6 years in China’s 10 agricultural provinces is as follows: the agricultural output value is about 231.9 billion Yuan, the number of labor is about 20.95 million persons, the land area is over 8348 thousand hectares, the mechanical force is over 49 million kilowatts, the fertilizer is about 75.49 million tons, and the use of the geomembrane is about 109.3 thousand tons.
Table 1. Data description.

| Variables       | Measurement | Mean     | S.d   |
|-----------------|-------------|----------|-------|
| Output ($y$)    | $10^8$ Yuan | 2,319.7  | 755.8 |
| Labor ($x_1$)   | $10^4$ Person| 2,095.2  | 2,100.4|
| Land ($x_2$)    | $10^3$ Hectare| 8,348.1  | 3,292.9|
| Mechanical ($x_3$)| $10^4$ Kilowatts| 4,950.4  | 3,751.6|
| Fertilizer ($x_4$)| $10^4$ Ton       | 7,548.6  | 21,882.1|
| Geomembrane ($x_5$)| $10^4$ Ton       | 10,9257.9| 83,368.3|
| Observation     |             | 60       |       |

4. Results and discussion

4.1. Maximum likelihood estimates (MLE) results

Table 2 shows the estimation results of the classical stochastic frontier model we found that labor ($\alpha_1$), fertilizers ($\alpha_4$) and geomembranes ($\alpha_5$) have a positive effect on production, with statistical significance, while the negative effects of land ($\alpha_2$) and mechanical power ($\alpha_3$) on agricultural production. For example, according to the result we know that the coefficient of the land is -0.1464, but the result is significant, which is contrary to the actual situation. The mechanical power is also inconsistent with the actual situation, the coefficient is 0.0023, but its result is not significant. In classical SFM, the maximum elasticity of fertilizer is 0.46, which means that a 1% increase in fertilizer allocated to crops will lead to an increase of 0.46% in agricultural output. This is similar to the findings of Ma et al. [5]. Therefore, it is necessary to appropriately increase the amount of fertilizer used.

Table 2. Maximum likelihood estimates of the stochastic frontier model.

| Parameter       | coefficient | Estimate | S.d   | z-value | Pr($-Z-$) |
|-----------------|-------------|----------|-------|---------|-----------|
| (Intercept)     | $\alpha_0$  | 7.5992   | 7.1295| 0.4274  | 16.680    | < 2.2e-16 *** |
| Inlabor         | $\alpha_1$  | 0.1423   | 0.1825| 0.0753  | 2.4234    | 0.015374 *   |
| Inland          | $\alpha_2$  | -0.1464  | -0.305| 0.0962  | -3.1704   | 0.001522 **  |
| Inmechanical    | $\alpha_3$  | 0.0023   | 0.0164| 0.0984  | 0.1670    | 0.867381     |
| Infertilizer    | $\alpha_4$  | 0.4081   | 0.4582| 0.1118  | 4.0972    | 4.181e-05 ***|
| Ingeomembrane   | $\alpha_5$  | 0.1839   | 0.2105| 0.0531  | 3.9648    | 7.347e-05 ***|
| sigmaSq         | $\sigma^2$  | 0.2012   | 0.0394| 0.0077  | 5.1276    | 2.934e-07 ***|
| gamma           | $\gamma$    | 0.0129   | 0.0002| 0.0721  | 0.0024    | 0.9981       |

4.2. Generalized maximum entropy (GME) results

To obtain the GME estimates, we select five support points for each regression coefficient, including one-sided inefficient components and two-sided random disturbances. Based on Greene’s assumption of linear uniformity using the cost function, and the results of the same data and function form, we expect that other parameters except the intercept are positive and less than one [13]. Therefore, we have chosen {-11, 5.5, 0, 5.5, 11} as support for intercepts. For other parameters, such as labor and land, we choose {-0.4, -0.2, 0, 0.2, 0.4} and {-0.5, -0.25, 0, 0.25, 0.5} as their support, other parameters are made in the same way. The value of coefficient is the value of $\alpha$, which corresponds to $\alpha_0, \alpha_1, \alpha_3, \alpha_4, \alpha_5$. The significance of the coefficient lies in whether the test result is significant. We say that when the coefficient value is between the lower value and the upper value, the result is optimal. From the results in table 3, it can be seen that the coefficient of the labor force is 0.138, which is significant between the lower limit of 0.0571 and the upper limit of 0.3296. At the same time, we also found that the land coefficient is negative, and the upper and lower limits include 0, which is not significant, indicating that it is consistent with the actual situation. See Table 3 for details.
Table 3. Generalized maximum entropy estimates of the stochastic frontier model.

| Parameter | coefficient | mean  | S.d   | Lower  | Upper  |
|-----------|-------------|-------|-------|--------|--------|
| (Intercept) | $\alpha_0$ | 7.6021 | 8.5606 | 1.0166 | 7.5774 | 10.043 |
| ln labor  | $\alpha_1$ | 0.1385 | 0.1571 | 0.088  | 0.0571 | 0.3296 |
| ln land   | $\alpha_2$ | -0.1252 | 0.1921 | 0.1765 | -0.1307 | 0.4756 |
| ln mechanical | $\alpha_3$ | 0.0029 | 0.0372 | 0.0422 | 0.0017 | 0.0806 |
| ln fertilizer | $\alpha_4$ | 0.3951 | 0.5386 | 0.1183 | 0.3809 | 0.7448 |
| ln geomembrane | $\alpha_5$ | 0.1772 | 0.2261 | 0.0571 | 0.1618 | 0.7448 |

4.3. Comparison of GME with MLE
Through the analysis of the above results. We find that the coefficient of land in classical SFM with MLE is -0.1464, but the result is significant, which is contrary to the actual situation. In the SFM with GME estimation, we find that the land coefficient is negative, although the result is between the upper limit value and the lower limit value, but this interval includes 0, which means that it is not significant and it is consistent with the actual situation. This reflects that GME is more suitable for technical efficiency analysis than MLE.

4.4. Estimation of technical efficiencies (TE)
Table 4 summarizes the technical efficiency of Shandong, Henan, Jiangsu, Hebei, Sichuan, Hubei, Hunan, Guangdong, Heilongjiang, and Liaoning using their generalized maximum entropy. The average technical efficiency of Heilongjiang is the lowest, which is 0.5486, while the average of Guangdong’s technical efficiency is the highest, which is 0.7293. In 2009, the minimum technical efficiency of Heilongjiang was the lowest, only 0.4173. The maximum technical efficiency of Guangdong is 0.8646, the maximum technical efficiency of Henan is only 0.6178. Even so, the difference in the maximum technical efficiency between the two provinces is not very large. It can be said that the maximum of these 10 provinces is very close. During the period from 2002 to 2017, the technical efficiency of the 10 provinces fluctuates between approximately 0.4 and 1. This means that each province has high agricultural productivity.

Table 4. Summary of efficiencies by top ten provinces.

| Province   | Mean  | S.d   | Min   | Max   |
|------------|-------|-------|-------|-------|
| Shandong   | 0.6074 | 0.0714 | 0.5021 | 0.6664 |
| Henan      | 0.5891 | 0.0339 | 0.5488 | 0.6178 |
| Jiangsu    | 0.6591 | 0.0855 | 0.6117 | 0.8236 |
| Hebei      | 0.5704 | 0.1413 | 0.4595 | 0.7753 |
| Sichuan    | 0.6705 | 0.1247 | 0.5731 | 0.8561 |
| Hubei      | 0.5931 | 0.1349 | 0.4292 | 0.8234 |
| Hunan      | 0.6029 | 0.0166 | 0.5838 | 0.6261 |
| Guangdong  | 0.7293 | 0.0761 | 0.6511 | 0.8646 |
| Heilongjiang | 0.5486 | 0.1333 | 0.4173 | 0.7692 |
| Liaoning   | 0.6357 | 0.1187 | 0.4388 | 0.7527 |

According to the annual average technical efficiency (TE) of the sample provinces, we selected the top three provinces in terms of annual average technical efficiency: Guangdong, Sichuan, Jiangsu and the three provinces with the bottom of the average annual technical efficiency: Henan, Hebei and Heilongjiang. The result is shown in figure 1. After six years of development, Guangdong Province has increased from 0.733 in 2007 to 0.865 in 2012, which is close to the production frontier. Even if the backward Henan, Hebei and Heilongjiang, from 2007 to 2012, the technical efficiency of the three provinces has approached or reached the average technical efficiency of China’s top ten agricultural
production provinces of 0.62. For example, the technical efficiency of Hebei Province in 2011 was as high as 0.775.

![Figure 1](image)

**Figure 1.** Annual technical efficiency of Guangdong, Sichuan, Jiangsu, Henan, Hebei and Heilongjiang.

5. **Conclusions**

This study uses the SFM with GME to analyze the technical efficiency of China’s top 10 agricultural provinces and compares it with the classical SFM with MLE. A notable feature of GME is that researchers can impose a priori information on the estimated parameter, this function is particularly suitable for applications in the field of effective boundary estimation. GME combines the advantages of SFA and MLE. Therefore, it is suggested that GME analysis can be used more frequently in future agricultural technology efficiency analysis in China. The results show that: (1) China’s agricultural production technology efficiency has increased year by year, and advanced agricultural areas have approached the frontiers of technical efficiency; (2) Since the average technical efficiency of all samples in the GME is 0.62, China’s technological efficiency still has much room for improvement. Finally, it is proposed that: (1) In the future, the research on agricultural technical efficiency can consider the generalized maximum entropy analysis method; (2) From the perspective of improving the technical efficiency, China should increase investment in fertilizers and geomembrane, because...
these factors can directly stimulate agricultural output. In particular, the provinces in underdeveloped areas must increase investment. At the same time, mutual assistance and learning between provinces is also a good way to improve the efficiency of agricultural technology.

References
[1] Aigner D J, Lovell C A K and Schmidt P 1977 Formulation and estimation of stochastic frontier production function models *Econometrics* 6(1) 21-37
[2] Meeusen W and van den Broeck J 1977 Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error *International Economic Review* 18(2) 435-444
[3] Tian W and Li M 2009 Analysis on the Technical Efficiency of Agricultural Production in China based on SFA—Taking Rape Production as an Example *Research on Productivity* 21 55-73
[4] Liu J X, Sanzidur R, Songsak S and Aree W 2017 Enhancing Productivity and Resource Conservation by Eliminating Inefficiency of Thai Rice Farmers: A Zero Inefficiency Stochastic Frontier Approach *Sustainability* 9(5) 770
[5] Ma J, Liu J and Songsak S 2018 Technical efficiency analysis of China’s agricultural Industry: a stochastic frontier model with panel data *Predictive Econometrics and Big Data* pp 454-463
[6] Sengupta J K 1992 The Maximum Entropy Approach in Production Frontier Estimation *Mathematical Social Sciences* 25(1) 41-57
[7] Zhang X and Fan S 2001 Estimating Crop-Specific Production Technologies in Chinese Agriculture: A Generalized Maximum Entropy Approach *American Journal of Agricultural Economics* 83(2) 378-388
[8] Rezek J P and Campbell R 2007 Cost Estimates for Multiple Pollutants: a Maximum Entropy Approach *Energy Economics* 29(3) 503-519
[9] Campbell R, Rogers K and Rezek J 2008 Efficient Frontier Estimation: a Maximum Entropy Approach *Journal of Productivity Analysis* 30(3) 213-221
[10] Macedo P and Silva E 2010 A Stochastic Production Frontier model with a Translog Specification using the Generalized Maximum Entropy Estimator *Economics Bulletin* 30(1) 587-596
[11] Battese G E and Corra G S 1977 Estimation of a production frontier model: with application to the pastoral zone off eastern Australia *Aust. J. Agric. Econ.* 21(3) 169-179
[12] Golan A, Judge G and Miller D 1996 *Maximum Entropy Econometrics: robust estimation with limited data* (Wiley)
[13] Greene W H 1990 A Gamma-distributed Stochastic Frontier Model *Journal of Econometrics* 46(1/2) 141-164