The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Plot size and maize production efficiency in China: agricultural involution and mechanization

Yufeng Luo, University of Georgia, E-mail: lyf@uga.edu

Selected Paper prepared for presentation at the 2018 Agricultural & Applied Economics Association Annual Meeting, Washington, D.C., August 5-August 7

Copyright 2018 by Yufeng Luo. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract: There has been a lot of debates on the relationship between land fragmentation and agricultural productivity, and there is still no decisive conclusion. Using a unique dataset with a large range of plot size, this paper examines the relationship between land fragmentation and production efficiency in China’s maize production. The efficiency is both being measured by the productivity (yield under given inputs) and net profit of unit land. Contrast with most previous studies where a linear or quadratic linear form being imposed before empirical estimation, this paper uses a semi-parametric method without assuming a specific relationship in empirical analysis. The results show that small size plots have higher land productivity but lower profit due to the intensive use of inputs; big size of plots have higher land productivity and higher profit owing to mechanization and less labor inputs. But, there is a large range of medium size where yield keeps in a low level with little change (while profit keeps increasing), and those large range cannot be find based on traditional linear or quadratic linear models. Our findings imply that China’s land productivity will witness a significant decline during the process of land consolidation, which may have negative effects on its grain security, and only after the size of plots reach a much higher level, land productivity will start to increase. On contrary, what farmers most care about, net profit of unit land, will always have upward trend during the process of land consolidation.
Keywords: Land fragmentation, plot size, production efficiency, semi-parametric model

1. Introduction

In the recent decade, China’s agricultural system is undergoing an apparently transformation from small-scale farms based on “Household Contract Responsibility” to modern family farms. Urbanization and industrialization during past two decades have provided Chinese farmers with much more non-agricultural job choices. By the end of 2017, there are more than 286 million rural labors worked on urban and nonagricultural sectors (National Bureau of Statistics of China, 2018). Correspondingly, Average hours spent on the farming activity per household has declined from 3,500 hours in 1991 to 1,400 hours in 2009 (de Brauw et al., 2013). Under the background that a huge number of farmers has been attracted to off-farm sector with high wage rates, land fragmentation gradually becomes a constrain of the development of agricultural productivity because of the decreasing intensity of labor input.

To promote efficient use of land resource and boost agricultural productivity, China central government has implemented subsidy and other incentive programs to encourage the expansion of farm sizes (Gale, 2013; Rada et al., 2015). China’s “No. 1 Central Document” of 2013 first emphasizes the development of large farms, also known as “new family farms” as a strategy for agricultural growth and national food security. In addition, land rental markets as well as land transfer service centers in rural China were established to assist land transfers among farmers and to consolidate land to achieve the economy of scale in agricultural production (Jin and Deininger, 2009; Trappel, 2015). China’s government also provides more production subsidy to large farmers than the small size farmers, to encourage them expanding farm size (Gale, 2013).

Grain production plot size in China is getting much larger than before due to the combine of internal motivation and external policy thrust. Does it help to improve the production efficiency in China’s agricultural production and improve China’s grain security?

Theoretically, it’s possible that taking into account of the adoption of new technology, such as big machines, land consolidation may help to achieve a higher productivity. However, on the other side, with the increase of plot size, the labor inputs (and possibly other inputs) on unit of land may decline, and changed the previously intensive farming system to an extensive system which may have negative impacts on land productivity. The existing arguments favoring positive relationship between plot size and land productivity mostly rely on the assumption that large scale farmers are more willing to adopt new technologies and mechanization. But, some studies also show that the adoption of mechanization and new technology only after the plot size reach a threshold. Those argument also support a U shape relationship between plot size and land productivity.

The goal of this paper is to empirically test the relationship between plot size and production efficiency of farmers planting maize in China. Based on a set of pooled cross section data (2012 and 2015) of 1845 plots from 9 provinces in China, this paper uses semi-parametric model to estimate the relationship and figure out policy implications from the findings. This model relaxes the linear or quadratic linear assumption between plot size and production efficiency in most existing studies. Our findings imply that the relationship between plot size and land productivity seems like a “pan” with a long flat bottom. One interesting finding is that, even though the land productivity will witness a significant decline during the process of land consolidation, farmers still have the incentives to increase plot size due to the continuous increasing average profit per unit of land. So, there seems a conflict between national food security and farmers’ income, since with land consolidation, the productivity may decline while farmers’ income from both agricultural and nonagricultural sectors may increase. Our results implies that, if China’s government want to ensure food security and improve production productivity, subsidize on the adoption of new technology and mechanization of medium size farmers or the outsourcing service in agricultural production may be a better choice than just promoting the plot size.

The rest part of this paper is organized as follow. In the next section, we analyze related literature and states our contributions to the literature. In section 3, we discuss the data and provide some descriptive statistics, followed by the introduction of empirical econometric model. In section 4, we present the results of models and robust checking. In section 5, we conclude by drawing out policy implications.

2. Literature Review

The debate on the relationship between land fragmentation and land productivity has lasted over decades since Chayanov (1926) and Sen (1962) firstly found that small farms are more productive than large farms in Russia and India. Since then many empirical studies have been conducted on this topic. From the very beginning, ‘inverse relationship’ seems to be the stylized results (Berry and Cline, 1979; Bhalla, 1979; Rao and Chotigeat, 1981; Carter, 1984; Cornia, 1985). Though this relationship has been obviously attenuated during the Green Evolution period (early 1970s), in which big farms benefit more from new technologies introduced (Deolalikar, 1981). The inverse relationship remains an important empirical finding in most studies post-dates the Green Revolution (Verschelde et al. 2013). Henderson (2015) found an overall non-linear relationship with ‘inverse relationship’ under 38 hectares, when adding quadratic component to the linear model. However, studies using data from China and some African countries indicates that ‘inverse relationship’ doesn’t apply to these areas (Wang et al., 2015; Dorward, 1999). There are also studies show that the negative relationship between plot size and productivity does not exist. For example, Deininger et al. (2006) using dataset from Albania finds no support for the argument that fragmentation reduces productivity. There are four main reasons are put forward to explain the potential relationship between plot size and land productivity, though none of them can explain the problem thoroughly by itself.

A first reason is the heterogeneity between small farms and large ones. Plot size and productivity relationship (PPR) showed by existing studies was observed in household level, hence explanations related to the heterogeneity between small farms and large ones prevailed for a long period of time. In this branch, imperfect factor market is the most commonly suggested explanation for PPR. For example, the imperfect labor marked differentiate the labor input between small farms and big farms. Many researchers noticed that the relationship between labor intensity and farm size is also inverse (Cater, 1984). One easy way to link these two factors is that imperfect labor market leaves some labors cannot find a job in market. Small farms with these underemployed labors thus input the surplus labors into their family plots to achieve high productivity while big farms rarely suffer from this problem. Ali (2015) found that the 'inverse relationship' disappears if the wage of family member is valued at market price. Heltberg (1998) made a supplement that imperfect labor market cannot alone be sufficient to cause 'inverse relationship' and put forward that multiply imperfect factor markets (such as land market and credit market) should be taken into account altogether. This analysis framework makes sense because small farmers can adjust land (or other factors) input to avoid excess labor input in per unit area of land if only labor markets are imperfect. Meanwhile, from a different perspective in this branch of explanation, another possible heterogeneity is farm skills. Assuncao et al. (2003) claim that skilled peasants have more tendency to work in their own small family lands and this self-selection problem may also be the source of "inverse relationship".

However, there exists study finding that the "Inverse relationship" is not only observed in household level but also in plot level (Assuncao et al., 2007). This finding indirectly suggests that cross-household heterogeneity (such as difference in labor input quantity and quality because of imperfect labor markets) does not suffice to explain the relationship.

A second explanation of the "Inverse relationship" is related to omitted variables. Some researchers (Bhalla and Roy, 1988; Benjamin, 1995) pay their main attention to the missing variables, the most popular one of which is land quality. This hypothesis was firstly proposed by Sen (1975), who noticed that 'inverse relationship' was more pronounced between regions than within regions. Bahalla and Roy (1988) found that 'inverse relationship' were less pronounced when using geographical disaggregation approach. Besides, there were other studies using instrument variables to address missing land quality problem and found 'inverse relationship' vanished after adjustment (Benjamin, 1995; Chen, 2011). With the upgraded data sets achieved later, many studies (Verschelde et al., 2013; Ali, 2015; Desiere et al., 2018) can control the land quality (either self-reported by farmers or measured by soil test) directly in their model. Most of these studies came to the similar conclusion: land quality can account for part but not all of the 'inverse relationship'.

The third alternative explanation emerges after 2000: measurement error. Lamb (2003) compared this potential explanation with prevalent explanations discussed above and concluded that measurement error may play an important role in explanation of PPR. To solve the measurement error problem, some resent studies used new dataset applying advanced measurements to fix the potential measurement errors. Desiere et al. (2018) estimated the yield based on crop cuts instead of self-reported production by farmers, while others estimated the land size with GPS measurement instead of self-reported land size by farmers (Carletto et al., 2013; Carletto et al., 2015). However, the conclusion is not consistent. Desiere et al. (2018) put forward that “Inverse relationship” disappears after adjustment on production data, while Carletto et al. (2013) found that “Inverse relationship” were even strengthened after using more accurate data on land area.

Last but not the least explanation is the scale range and the methodological issue. The empirical studies using cross-section data and old dataset always assume the relationship is linear for convenience (Carter, 1984; Cornia, 1985). This assumption may hold true if the scale range is not big enough but may fail if we study on a bigger range of scales. As Lipton (2010) mentioned, we cannot view big farms as the simple replicas of small farms because technology applied varies during the expansion of plot size. Thus, the relationship between land size and land productivity may not be linear. To correct this issue, some other researchers (Kimhi, 2006; Henderson, 2015) assume that the form of key independent variable (such as farm size or operated area) can be nonlinear, but they still estimate the effect in linear parametric model. Verschelde et al. (2013) firstly applied non-linear model to measure the relationship between farm size and output, and his results show that the “Inverse relationship” cannot be rejected, but one potential drawback of his study is the small range of farm size in his sample.

This paper analyze the relationship between plot size and production efficiency with semi-parametric model. With this method, we allow any non-linear relationship between plot size and production efficiency and only assumes the effect of other control variables on production efficiency are linear. Furthermore, our rich dataset allows us to have a full view of the relationship because the scale range in our sample is big enough (ranging from 0.1 to 243 mu). Another contribution of this study is that we analyze the relationship between plot size and productivity based on plot data collected from different households, with this data, we can control the potential effects of plot characteristics such as land fertility, and can also control the potential effect of heterogeneity among different households.

3. Data and Model Specification

3.1 Data

The data used for this paper were collected from 9 provinces of China (Shandong, Shaanxi, Jilin, Zhejiang, Henan, Gansu, Hunan, Sichuan, Heilongjiang). The total cultivated areas of maize in these provinces account for more than half of the total maize areas in China (NSBC 2017). A multi-stage and random sampling methods were applied for sample selection. In 2012, 24 counties of 8 provinces were randomly selected. Then two townships in each county and two villages in each township were selected based on their average farm sizes. In each village, we randomly chose 12 households. In each household, two maize plots in each maize-production household were randomly selected, and, one plot will be surveyed if there is just one maize plot in the household. Of the households surveyed, not all were involved in maize production. Therefore, 947 maize plots of 582 households were contained in the survey carried out in 2012 (217 households with one maize plot and 365 households with two maize plots).

In order to expand the range of plot size covered by the samples, we jointly used the data from another survey conducted in Heilongjiang, Henan, Sichuan and Zhejiang Province in 2014, which is more focused on large scale farming. Using a stratified random sampling procedure, four counties in each province, two townships in each county and four villages in each township were randomly selected. In each village, all the households with a large farm size (more than three times of local average farm scale) were surveyed, and if there were more than 12 households qualified with this criterion, 12 households were randomly surveyed. We also randomly selected 1 or 2 households with normal farm size around each large-farm-size household. Finally, two plots were selected in each household. In total, 1040 households were surveyed. Since not all the households surveyed were involved in maize production, 898 maize plots were surveyed in 2014.

We gathered detailed information on households' agricultural production at plot level in 2012 and 2014, including output quantities, input quantities of all agricultural factors in each production links and characteristics of plots. We also asked farmers whether their agricultural production had suffered from any natural disasters that contributed to yield reduction. Demographic information about the households, such as family size, education, age and gender of household head, were also collected. Our study eventually used 1845 plot-level observations.

3.2 Descriptive analysis

In this study, we use yield, net profit per mu as the measurement for production efficiency. Yield has been standardized to 14% water content of corn. As the data provides detailed information on input quantities in each production process and local prices of input factors, we aggregate values of all inputs and calculate to get costs. Net profit is calculated by total output value minus input costs. To make it comparable for the two years, all prices being used in 2014 have been adjusted to the constant price of 2012 using provincial Agricultural input and output price index.

We use plot size as the main explanatory variable. Following the existing literatures (Lamb, 2003; Kimhi, 2006; Desiere et al., 2018), major factors affecting production efficiency being considered are production inputs, natural disasters, land features and household characteristics. Studies of Assuncao (2007) and Barrett (2010) show that the quantity of inputs are endogeneity, so we do not include production inputs in the model as control variables. A dummy variable about whether the plot suffered from natural disasters that caused yield reduction represents production risks. Other control variables reflecting the heterogeneity of land conditions include land property using right, land slope, land quality, soil type, distance from land to home, dummy variable representing the irrigation condition as well as number of land plots sowed by the household. Household's characteristics include family size, the share of labors in the family, education, gender, age and social status (measured by whether household head is a village officials or not) of household head. In addition to the above factors, province dummies are introduced to capture the potential regional heterogeneity. Table 1 provides a description and sample statistics for plot-level variables used in the analysis.

Table1 Variable description and summary statistics

Variable	Description	Mean	SD	Min	Max
Dependent variables					

Yield	Yield of maize(0.5kg/mu)	958.3	301.4	93.02	1970
Profit	Profit of maize(yuan/mu)	380.2	437.7	-1400	1821
Independent variables					
Key variable					
Plot size	Area of maize plot(mu)	9.317	18.31	0.100	243
Risk of yield reduction					
Disaster	Whether suffered from any natural disasters(1=Yes,0=No)	0.428	0.495	0	1
Land features					
Property	Property of land(1=Owned by household,0=Rented)	0.267	0.442	0	1
Land slope	Land slop(1=Plain,0=Others)	0.753	0.431	0	1
Land quality	Self-report land quality(1=Good, 2=Medium, 3=Bad)	1.713	0.702	1	3
Irrigation	Whether land can be irrigated(1=Yes,0=No)	0.582	0.493	0	1
Soil type	Soil type (1=Sand oil, 0=Others)	0.240	0.427	0	1
Distance	Distance form land to home(0.5km)	1.740	3.491	0	100
Plots number	Number of land parcels (No.)	6.786	12.12	1	300
Household characteristics					
Education	Schooling years of household head(years)	7.264	2.921	0	16
Gender	Gender of household head(1=Male,0=Female)	0.940	0.309	0	6
Age	Age of household head(years)	51.87	10.53	24	85
Leader	Whether household head is a village official or not(1=Yes, 0=No)	0.247	0.431	0	1
Family size	Number of family members(No.)	4.623	1.811	1	15
Labor rate	Share of labors (%)	0.700	0.219	0	1
Time dummy					
Year	Year of agricultural production(0=2012, 1=2014)	0.487	0.500	0	1

Source: Author's own calculations based on surveyed data

3.3 Econometric models

Most existing literatures studied the relationship between plot size and production efficiency based on a linear parametric models, which will fail to capture the potential non-linear relationship between production efficiency and plot size (Verschelde et al., 2013). In order to relax constraint, we apply a plot level semi-parametric model specified as following:

$$Y_{ij} = \beta_0 + g(z_{ij}) + \beta_1 x_j + \beta_2 x_{ij} + \varepsilon_{ij} \quad (1)$$

$$P_{ij} = \beta_0 + g(z_{ij}) + \beta_1 x_j + \beta_2 x_{ij} + \varepsilon_{ij} \quad (2)$$

where Y_{ij} is yield in plot i of household j, P_{ij} is net profit in plot i of household j. z_{ij} denotes the plot size of household j. $g(z_{ij})$ measures the non-parametric relationship between farm size and productivity. x_{ij} is a set of control variables in plot level, including yield reduction due to natural disaster and land features. x_j is a set of control variables at household level, and ε_{ij} is an error term accounting for idiosyncratic shocks to the dependent variables. The plot level specification is used as it allows us to better control for heterogeneity across plots.

Estimation of a semi-parametric model is generally performed in the following two steps. First step is to estimate the parameters β_1 and β_2 then apply nonparametric methods to recover the unknown function. Take the conditional expectation of both sides of (1) with respect to z. This leads to

$$\mathrm{E}(Y_{ij}|z_{ij}) = \beta_0 + g(z_{ij}) + \beta_1 E(x_j|z_{ij}) + \beta_2 E(x_{ij}|z_{ij}) \tag{3}$$

By subtracting (4) from (1), we have

$$Y_{ij} - \mathrm{E}(Y_{ij}|z_{ij}) = \beta_1[x_j - E(x_j|z_{ij})] + \beta_2[x_{ij} - E(x_{ij}|z_{ij})] + \varepsilon_{ij} \tag{4}$$

Follow Robinson's suggestion, use local-constant least-squares (LCLS), we can estimate each conditional mean, $\mathrm{E}(\widehat{Y_{ij}|z_{ij}})$, $E(\widehat{x_j|z_{ij}})$ and $E(\widehat{x_{ij}|z_{ij}})$. Then the estimators of parameters, $\widehat{\beta_1}$ and $\widehat{\beta_2}$ can be obtained by ordinary least squares (OLS) regression of $Y_{ij} - \mathrm{E}(\widehat{Y_{ij}|z_{ij}})$ on $x_j - E(\widehat{x_j|z_{ij}})$ and $x_{ij} - E(\widehat{x_{ij}|z_{ij}})$.

For the nonparametric component, replace β_1 and β_2 with $\widehat{\beta_1}$ and $\widehat{\beta_2}$ back in (1), we have

$$Y_{ij} - \widehat{\beta_1}x_j - \widehat{\beta_2}x_{ij} = \beta_0 + g(z_{ij}) + \varepsilon_{ij} \tag{5}$$

Use kernel regression, the estimator of $g(z_{ij})$ and its gradient vector are obtained as

$$\begin{pmatrix} \hat{g}(z) \\ \frac{\partial \hat{g}(z)}{\partial z} \end{pmatrix} = [\textstyle\sum_{ij} K_h(z_{ij}, z) \begin{pmatrix} 1 \\ z_{ij}-z \end{pmatrix}]^{-1} \sum_{ij} K_h(z_{ij}, z) \begin{pmatrix} 1 \\ z_{ij}-z \end{pmatrix} (Y_{ij} - \widehat{\beta_1}x_j - \widehat{\beta_2}x_{ij}) \tag{6}$$

where $K_h(z_{ij}, z)$ is the kernel weighting matrix. In this analysis, we chose Gaussian kernel function. Specially, we are interested in the effects of a change in z on estimated Y, known as the partial effects (or marginal effects) $\partial\widehat{\mathrm{Y}}/\partial z$. Based on equation (6), we can calculate the partial effect at every z_{ij}. Using the same procedure, we can also estimate $\partial\widehat{\mathrm{P}}/\partial \mathrm{z}$ and $\partial\widehat{\mathrm{C}}/\partial \mathrm{z}$.

To test whether nonparametric function can be approximated by some parametric polynomial alternative, we apply Hardle and Mammen's (1993) test, which compares the nonparametric and parametric regression fits using squared deviations between them. The test-statistic is:

$$T_n = N\sqrt{h}\textstyle\sum_{ij}[\hat{g}(z_{ij}) - \hat{g}(z_{ij}, \theta)]^2 \pi(.) \tag{7}$$

where $\hat{g}(z_{ij})$ is the nonparametric function estimated in (6), $\hat{g}(z_{ij}, \theta)$ is an estimated parametric function, N is the number of observations, h is the bandwidth used and $\pi(.)$ is a weighting function for the squared deviations between fits. Following Hardle and Mammen's suggestion, we use wild bootstrap to obtain the simulated values of test statistics. A rejection of the null (H0: $\hat{g}\left(z_{ij}\right) = \hat{g}\left(z_{ij}, \theta\right)$) means that the estimated nonparametric function is significant different from the parametric polynomial tested.

4. Results and robustness test

4.1 Results on plot size and yield from semi-parametric estimation

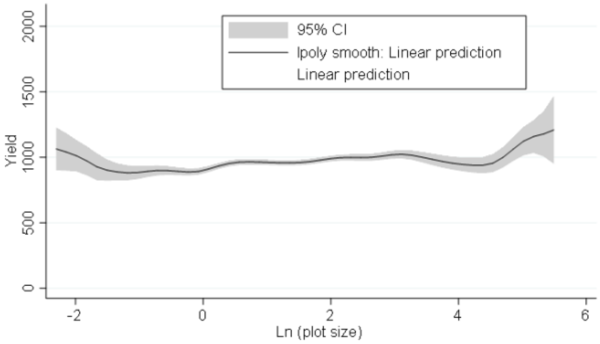

Fig. 1 Non-parametric fit of yield with plot size

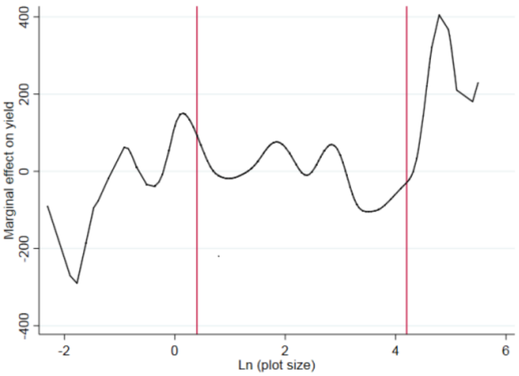

Fig. 2 Marginal effects of plot size on yield

We illustrate the non-parametric results by showing directly the estimated of dependent variable as a function of the value of independent variable, keeping other parametric variables controlled. In addition, 95% confidence intervals are visualized around the nonparametric fit. A significantly increasing (decreasing) curve illustrates a significant positive (negative) effect of the regressor on dependent variable. To be more intuitively, marginal effects of plot size on dependent variable is also reported.

As far as the effects of other variables are considered, Figure 1 shows that the relationship between plot size (in log) and yield seems like a “pan” with a long flat bottom. Indeed, as the plot size increases, yield first deceases, then witnesses a long stable period (the average effect of plot size to yield close to zero), and increases significantly when plot size exceeds a certain large value. Obviously, the relationships between plot size and yield undergo three stages, which demonstrates that a common specification cannot explain this process. Therefore, we divide samples into three groups according to Figure 1. The first one includes samples with small plot size, which shows a negative effects of plot size on yield. The second one covering medium plot size represents the flat bottom of the “pan” relationship. And the last one is the stage of increase relationship.

Figure 2 shows the marginal effects of plot size on yield obtained from semi-parametric estimation. For small size, the marginal effects are far less than zero at beginning, and then rise slightly above zero. For medium size, the marginal effects fluctuate around zero. For large size, the curve goes up sharply, the marginal effects fluctuate mainly above 0, which shows a positive relationship between plot size and productivity during this stage. The results of Figure 2 is generally consist with Figure 1. When we check if the quadratic approximation is appropriate, it turns out that this assumption is clearly rejected by Hardle and Mammen’s (1993) test (see Table 3 Colum (1)). It means that we cannot simply use a pure quadratic parametric model to fit the relationship between plot size and yield.

The first column of Table 2 presents the parametric results from semi-parametric estimation on yield. The results show that suffering from disasters has a significant negative impact on yield. While good land characteristics (such as good-quality, plain and irrigable) have significant positive impacts on yield, which confirms the findings of Benjamin (1995) and Chen (2011). Moreover, yield is also positively influenced by the education level and age of household head. And if the household head is a leader in a village, the household is more likely to get a higher yield. What's more, compared to 2012, yield of 2014 is significantly higher.

4.2 Results on plot size and profit from semi-parametric estimation

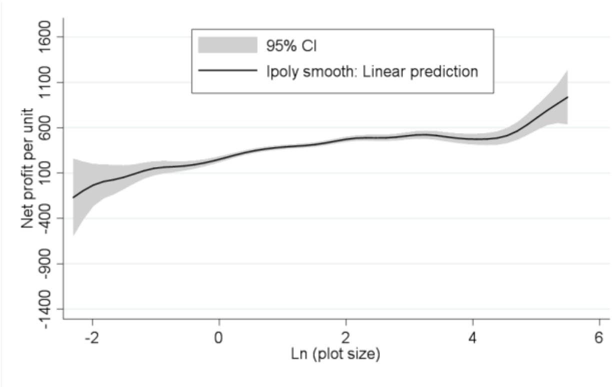

Fig. 3 Non-parametric fit of profit with plot size

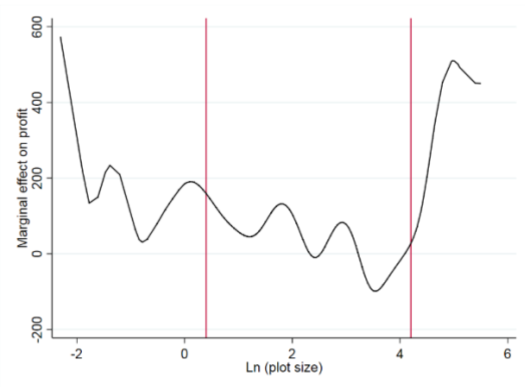

Fig. 4 Marginal effects of plot size on profit

In Figure 3 the relationship of plot size and net profit per mu is shown. We find that profit is monotonically increasing with plot size, which indicates that the farmers with all plot sizes have incentives to increase plot size due to the continuous increasing average profit per unit. However, from small size to medium size, the growth rate of net profit decreases. And this rate increases sharply when it comes to large size (Figure 4). Also, when we use Hardle and Mammen's (1993) method to test the difference between nonparametric and linear regression, the null is rejected (see Table 3 Colum (2)) which means that a pure linear parametric model with a common coefficient could not be used to estimate the effect of plot size on profit per unit.

The results of the parametric parts (see Table 2 column (2)) show that, similar to the effects of plot size on yield, suffering from natural disasters and good land characteristics have significant negative and positive impact on profit per unit. Except being a leader in a village, other characteristics of household head have no significant effects on profit per unit. In addition, share of household labors affects profit per unit negatively and significantly. Because there may not be enough off-farm jobs available if the labor market functions inefficiently. Thus, surplus of labors were put into agricultural production due to the absence of off-farm jobs. Moreover, even though the average yield in 2014 is significantly higher than that of 2012, the average profit per unit in 2014 is much lower than that of 2012. Because the growth rate of maize output value is much lower than that of input values. To be more specific, maize price was stable between 2012 and 2014 due to China's corn stockpiling policy, while the price of agricultural inputs increased sharply during this time (NSBC, 2017).

4.3 A conflict between food security and farmers' income

As the regression results shown above, we have two main findings. First, compared to medium size plots, small size plots have higher average yield but lower average net profit, while large size plots have both higher average yield and higher average net profit. Second, there is a large range of medium size plots that witness lower and stable yields. And its growth rate of net profit is lower than small size plots' and large size plots'. From the perspective of food security, China's grain productivity may decline during the process of land consolidation, especially from small size plots to medium size plots. However, from the perspective of farmers' income, all farmers will earn more due to the consolidation of land. Therefore, there seems a conflict between national food security and farmers' income in short term. And what cause this conflict are crucial.

On one hand, yield declines at the beginning of the land consolidation, and there is no significant increase for a large range of medium size, which may have a negative impact on China's food security. The higher yield of small size plots compared with medium size plots is mainly caused by the intensive use of inputs. The average cost per unit of medium size is 536.1 yuan/mu, while that of small size is 730.4 yuan/mu, 36.2% more than medium size. However, when the plot increases to a certain large size, large machinery which is most likely indivisible can be introduced. And the yield will increase due to the use of such technology (Otsuka et al., 2016). As presented in Table 4, the average machinery input per unit increases from medium size to large size. However, average machinery input per unit may underestimate the technological advances, because it fails to remove the effect of scale economics.

Table4 Means of input factors

Variable(per mu)	Total(mean)	Small size	Medium size	Large size
Average Cost	582.4	730.4	536.1	442.1
Machine input	95.08	75.44	101.3	111.7
Labor input	209.7	332.8	170.7	113.7
Self-labor input	193.7	325.7	152.3	75.56
Hired-labor input	16.04	7.117	18.41	38.09
Seed	58.59	64.18	56.82	54.39
Chemical	36.60	45.15	34.02	25.79
Fertilizer	177.5	201.1	170.7	136.5
Other input	4.843	11.77	2.626	0
Observations	1845	459	1345	41

Source: Author's own calculations based on surveyed data

On the other hand, all farmers' income increases during the process of land consolidation, which confirms to the research results of Otsuka et al. (2016). As the development of urbanization and technological advances in farm machinery in China, the price ratio of labor to machine use is increasing. And the efficient production method changes from labor-intensive to labor-saving and machine-using method (Otsuka et al., 2016). During this process, farmers tend to increase plot size, so as to substitute increasing costly labor by large machinery and realize scale economies. If the land rent market and labor market function effectively and smoothly, coupled with complementary and indivisible machines, only large size with high profit survive. In practice, however, it is likely that land and labor market do not function perfectly. So, labor-intensive small size and labor-saving large size may be observed at the same time. In other words, a positive relationship between plot size and net profit per unit may be observed.

Our paper confirms to this mechanism. For medium size, the increasing profit mainly results from the decrease of cost per unit, especially self-labor input. While, for large size, the increase of profit per unit owes to both the increase of yield and the reduction of cost. Figure 5 shows that there is an inverse relationship between plot size and cost per unit. In addition to machinery input, all of other inputs decrease from small size to large size (see Table 4), which means that compared with small size, both medium and large size have an advantage of low cost. Nevertheless, the decrease of cost is not just the proportional decrease of all input factors, but the more optimal allocation of resources. It is obvious that as the plot size increases, machinery input substitutes for labor input gradually (see Figure 5).

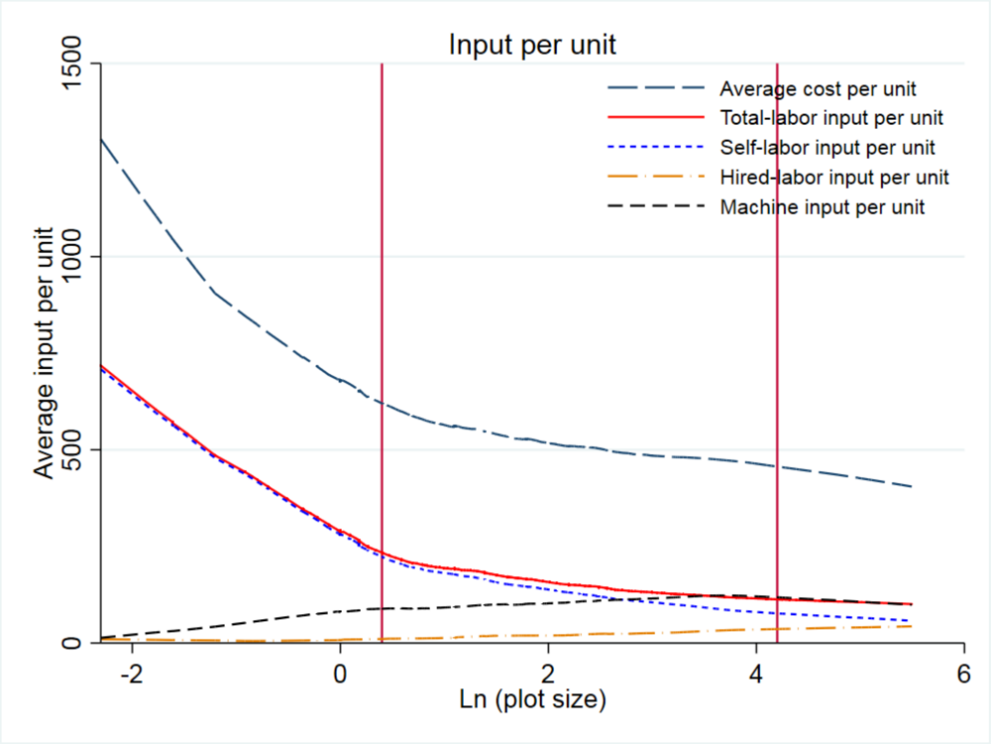

Fig. 5 Relationships between plot size and main input factors

There are two reasons may explain this result. First, as the increase of plot size, farmers have a larger demand for agricultural factors, which makes them have a stronger voice when bargaining with manufacturers. Second, according to the theory of "agricultural involution", farmers with smaller land size will input more labors into per unit of land even after marginal return of labor starts to decrease, which contributes to excessive inputs on per unit land. Inversely, farmers with larger land size can effectively eliminate the excessive inputs due to the limited household labors and the expanded land area.

4.4 Robustness test

To test the robustness of semi-parametric results, we apply a parametric estimation. Following the non-parametric curve between plot size and yield, we introduce dummy variables to classify samples into three groups (D_1 indicates small size which plot size is less than 1.5 mu. D_2 indicates medium size which plot size ranges from 1.5 to 65 mu, and D_3 indicates large size which plot size is larger than 65 mu). The parameter models can specified as

$$Y_{ij} = \alpha_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 cross_1 + \alpha_5 cross_2 + \alpha_6 cross_3 + \theta_j x_j + \theta_{ij} x_{ij} + \varepsilon_{ij} \quad (8)$$

$$P_{ij} = \alpha_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 cross_1 + \alpha_5 cross_2 + \alpha_6 cross_3 + \theta_j x_j + \theta_{ij} x_{ij} + \varepsilon_{ij} \quad (9)$$

where Y_{ij}, P_{ij}, x_j and x_{ij} are the same as those of equation (1) and (2). $cross_1$, $cross_2$ and $cross_3$ are the cross term of D_1, D_2, D_3 and plot size (in log), respectively. This specification assumes that each group has a separate coefficient. For small size ($D_1 = 1$ and $D_2, D_3 = 0$), the marginal effect of plot size on independent variable is α_4. Similarly, α_5, α_6 are the marginal effects of medium and large size.

Column (2) (4) in Table 5 present estimation results of equation (8) and (9). In order to compare the results of these piece-wise linear models with traditional linear or quadratic linear models, a quadratic linear specification on yield and a linear specification on profit are also estimated (see column (1) and (3)). Column (2) shows that large plot size has a significantly positive effect on yield, while medium size has no significant impact. For small size, the coefficient is negative but not significant, which consists with the wide confidence bands around the non-parametric curve of small size. Similar to the results of semi-parametric model, plot size has a significantly positive impact on profit per unit, but the coefficients of three groups are quite different (see column (4)). This means that, confirming to the Hardle and Mammen's test, a common specification for the whole sample is inappropriate. Furthermore, it is obvious that traditional linear or quadratic linear models fail to capture the real relationship between plot size and grain production efficiency.

Table5 Results from the parameter estimate

VARIABLES	(1) Yield	(2) Yield	(3) profit	(4) profit
Plot size	33.50***		79.67***	
	(3.083)		(8.222)	
(Plot size)^2	-3.940			
	(-1.503)			
D2		66.85***		89.22***
		(3.469)		(3.296)
D3		-997.0***		-1,428***
		(-2.845)		(-3.166)
cross1		-4.695		125.1***
		(-0.209)		(3.147)
cross2		7.946		50.70***
		(0.915)		(4.121)
cross3		236.3***		373.6***
		(3.152)		(3.894)

Disaster	-169.8***	-169.3***	-245.5***	-242.3***
	(-13.77)	(-13.84)	(-14.02)	(-13.97)
Property	-20.45	-18.38	-49.06**	-32.15*
	(-1.465)	(-1.328)	(-2.561)	(-1.650)
Land slope	102.1***	101.0***	110.5***	108.2***
	(7.084)	(7.058)	(5.509)	(5.394)
Land quality (basement=1,Good)				
Medium	-59.71***	-60.22***	-36.09**	-37.13**
	(-4.991)	(-5.046)	(-2.200)	(-2.277)
Bad	-108.2***	-106.9***	-110.2***	-108.8***
	(-6.190)	(-6.175)	(-4.073)	(-4.046)
Irrigation	75.46***	73.87***	77.37***	68.08***
	(5.392)	(5.284)	(3.858)	(3.365)
Soil type	-0.239	-1.730	-4.183	-6.895
	(-0.0190)	(-0.137)	(-0.221)	(-0.364)
Distance	-1.163	-1.054	-1.958	-1.388
	(-0.494)	(-0.457)	(-0.470)	(-0.360)
Plots number	-0.261	-0.180	0.222	0.214
	(-0.463)	(-0.325)	(0.485)	(0.473)
Education	5.055***	4.644**	2.986	2.271
	(2.676)	(2.473)	(1.052)	(0.803)
Gender	-32.87*	-32.26*	-38.03	-37.20
	(-1.958)	(-1.890)	(-1.615)	(-1.545)
Age	1.284**	1.192**	1.357	1.035
	(2.230)	(2.076)	(1.641)	(1.256)
Leader	49.62***	51.38***	65.98***	68.56***
	(4.085)	(4.245)	(3.904)	(4.075)
Family size	1.140	1.230	-0.817	0.687
	(0.385)	(0.415)	(-0.195)	(0.163)
Labor rate	-13.04	-8.213	-78.24**	-68.80**
	(-0.506)	(-0.319)	(-2.234)	(-1.983)
Year	140.1***	134.7***	-27.60	-45.71*
	(6.944)	(6.703)	(-1.063)	(-1.704)
Province Dummy	controlled	controlled	controlled	controlled
Constant	819.5***	793.3***	392.8***	369.3***
	(17.92)	(17.08)	(5.856)	(5.488)
Observations	1,845	1,845	1,845	1,845
R-squared	0.471	0.476	0.484	0.491
r2	0.471	0.476	0.484	0.491
ll	-12563	-12554	-13227	-13215

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4.5 Summary of the findings

Using a semi-parametric model and a unique dataset, we estimate the non-linear relationship between plot size and production efficiency. In this paper, we use yield, profit per unit as the indicators of production efficiency. As a result, there is a “pan” relationship between plot size and yield, and a positive relationship between plot size and profit per unit. This is significantly different from the relationship between plot size and productivity observed in the existing literature. For medium size, the increasing profit results from the decrease of cost per unit, especially self-labor input. While for large size, the increase of profit per unit owes to both the increase of yield and the reduction of cost. According to the “agricultural involution” theory, small farmers in China use excessive agricultural inputs. However, as plot size increase, excessive inputs per unit are eliminated gradually. What’s more, medium-size land witnesses a long period of yield trough. And yield increases with rising machinery input for large-size land. This means that mechanization improves yield significantly, however, yield of medium-size land is limited due to the indivisibility of large machinery and the lack of medium-size machinery.

Although there may be a negative effect on food security during the process of land consolidation, all farmers have incentives to increase plot size due to a strong positive plot size-profit relationship. Therefore, a transformation from small-scale farms to large-scale farms is an irreversible trend. Although yield keeps steady, medium-size plot can benefit from cost reduction. However, large-size plot can benefit from both yield increase and cost reduction. What’s more, larger plot size has more significant scale economies effects. In addition, high-quality and irrigable plot achieves higher yield and profit.

5. Conclusion and policy implication

In this paper, we use a semi-parametric yield function to curve the relationship between plot size and production efficiency, which allows for non-linear effects. And our rich dataset covered wide-scale plots allows us to have a full view of the relationship with soil and land quality controlled. Different from existing studies, we find a “pan” relationship between plot size and yield rather than the “inverse relationship”. This implies that large-size plots achieve higher yield compared with medium-size plots. Analyzing the change of agricultural inputs, mechanization and hiring skilled agricultural workers contribute to the yield increase. Also, land qualities matter. High perceived land quality is associated with high agricultural productivity. What’s more, even though the yield of maize will witness a long period of trough during the process of extension of land size, farmers still have the incentives to increase plot size due to the continuous increasing average profit per unit of land.

There are important policy implications that can be derived from the analysis in this paper. From a policy perspective, more efforts on new technology rather than on the increasing of plot size will be a better choice for Chinese government. Because farmers themselves have incentives to extend farm size given the positive relationship between plot size and profit. However, they will witness a significant yield decline during the initial expansion of production size. Therefore, we recommend improving access to modern technologies and services to break the “bottom” of medium-size plot. Especially, promote the research and use of medium-size machinery. Moreover, outsourcing services and agricultural trusteeship can create equal access for small and medium size farms to efficiency-improving technologies.

Reference

[1] Ali, Daniel Ayalew, and Klaus Deininger. "Is there a farm size–productivity relationship in African agriculture? Evidence from Rwanda." *Land Economics* 91.2 (2015): 317-343.

[2] Assunção, Juliano J., and Luis HB Braido. "Testing household-specific explanations for the inverse productivity relationship." *American journal of agricultural economics* 89.4 (2007): 980-990.

[3] Assunção, Juliano J., and Luis HB Braido. "Testing household-specific explanations for the inverse productivity relationship." *American journal of agricultural economics* 89.4 (2007): 980-990.

[4] Assuncao, Juliano J., and Maitreesh Ghatak. "Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity." *Economics Letters* 80.2 (2003): 189-194.

[5] Assuncao, Juliano J., and Maitreesh Ghatak. "Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity." *Economics Letters* 80.2 (2003): 189-194.

[6] Barrett, Christopher B., Marc F. Bellemare, and Janet Y. Hou. "Reconsidering conventional explanations of the inverse productivity–size relationship." *World Development* 38.1 (2010): 88-97.

[7] Benjamin, Dwayne. "Can unobserved land quality explain the inverse productivity relationship?." *Journal of Development Economics* 46.1 (1995): 51-84.

[8] Berry R A, Cline W R. "Agrarian structure and productivity in developing countries." *Population & Development Review*, 5.4 (1979):710.

[9] Bhalla, Surjit S. "Measurement errors and the permanent income hypothesis: evidence from rural India." *The American Economic Review* 69.3 (1979): 295-307.

[10] Bhalla, Surjit S., and Prannoy Roy. "Mis-specification in farm productivity analysis: the role of land quality." *Oxford Economic Papers* 40.1 (1988): 55-73.

[11] Bhalla, Surjit S., and Prannoy Roy. "Mis-specification in farm productivity analysis: the role of land quality." *Oxford Economic Papers* 40.1 (1988): 55-73.

[12] Carter, M.R., 1984. Identification of the inverse relationship between farm size and productivity: an empirical analysis of peasant agricultural production. Oxford Economic Papers, 36(1), pp.131-145.

[13] Carletto, Calogero, Sara Savastano, and Alberto Zezza. "Fact or artifact: The impact of measurement errors on the farm size–productivity relationship." *Journal of Development Economics* 103 (2013): 254-261.

[14] Carletto, Calogero, Sara Savastano, and Alberto Zezza. "Fact or artifact: The impact of measurement errors on the farm size–productivity relationship." *Journal of Development Economics* 103 (2013): 254-261.

[15] Carletto, Calogero, Sydney Gourlay, and Paul Winters. "From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis." *Journal of African Economies* 24.5 (2015): 593-628.

[16] Carletto, Calogero, Sydney Gourlay, and Paul Winters. "From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis." *Journal of African Economies* 24.5 (2015): 593-628.

[17] Carter, Michael R., and Keith D. Wiebe. "Access to capital and its impact on agrarian structure and productivity in Kenya." *American journal of agricultural economics* 72.5 (1990): 1146-1150.

[18] Chayanov, A.V. *AV Chayanov on the Theory of Peasant Economy*. Manchester University Press, Manchester, UK, 1926.

[19] Chen, Zhuo, Wallace E. Huffman, and Scott Rozelle. "Inverse relationship between productivity and farm size: the case of China." *Contemporary Economic Policy* 29.4 (2011): 580-592.

[20] Cornia, Giovanni Andrea. "Farm size, land yields and the agricultural production function: An analysis for fifteen developing countries." *World development* 13.4 (1985): 513-534.

[21] De Brauw, Alan, et al. "The feminisation of agriculture with Chinese characteristics." *The Journal of Development Studies*49.5 (2013): 689-704.

[22] Deolalikar, Anil B. "The inverse relationship between productivity and farm size: a test using regional data from India." *American Journal of Agricultural Economics* 63.2 (1981): 275-279.

[23] Desiere, Sam, and Dean Jolliffe. "Land productivity and plot size: Is measurement error driving the inverse relationship?" *Journal of Development Economics* 130 (2018): 84-98.

[24] Dorward, Andrew. "Farm size and productivity in Malawian smallholder agriculture." *The Journal of Development Studies*35.5 (1999): 141-161.

[25] Gale, H. Frederick. "Growth and evolution in China's agricultural support policies." *Economic Research Report* (2013).

[26] Hardle, Wolfgang, and Enno Mammen. "Comparing nonparametric versus parametric regression fits." *The Annals of Statistics* (1993): 1926-1947.

[27] Heltberg, Rasmus. "Rural market imperfections and the farm size—productivity relationship: Evidence from Pakistan." *World Development* 26.10 (1998): 1807-1826.

[28] Henderson, Heath. "Considering technical and allocative efficiency in the inverse farm size–productivity relationship." *Journal of agricultural economics* 66.2 (2015): 442-469.

[29] Henderson, Heath. "Considering technical and allocative efficiency in the inverse farm size–productivity relationship." *Journal of agricultural economics* 66.2 (2015): 442-469.

[30] Jin, Songqing, and Klaus Deininger. "Land rental markets in the process of rural structural transformation: Productivity and equity impacts from China." *Journal of Comparative Economics* 37.4 (2009): 629-646.

[31] Kimhi, Ayal. "Plot size and maize productivity in Zambia: is there an inverse relationship?" *Agricultural Economics* 35.1 (2006): 1-9.

[32] Lamb, Russell L. "Inverse productivity: Land quality, labor markets, and measurement error." *Journal of Development Economics* 71.1 (2003): 71-95.

[33] Lipton, Michael, and Richard Longhurst. *New seeds and poor people*. Taylor & Francis, 2010.

[34] NSBC (National Statistical Bureau of China). 2017. China Statistical Yearbook. China Statistic Press, Beijing, China. (in Chinese)

[35] Otsuka, Keijiro, Yanyan Liu, and Futoshi Yamauchi. "Growing advantage of large farms in Asia and its implications for global food security." *Global Food Security* 11 (2016): 5-10.

[36] Rada, Nicholas, Chenggang Wang, and Lijian Qin. "Subsidy or market reform? Rethinking China's farm consolidation strategy." *Food Policy* 57 (2015): 93-103.

[37] Rao, Vaman, and Tosporn Chotigeat. "The inverse relationship between size of land holdings and agricultural productivity." *American Journal of Agricultural Economics* 63.3 (1981): 571-574.

[38] Sen, Amartya K. "An aspect of Indian agriculture." *Economic Weekly* 14.4-6 (1962): 243-246.

[39] Trappel, René. *China's Agrarian Transition: Peasants, Property, and Politics*. Lexington Books, 2015.

[40] Verschelde, Marijn, et al. "Challenging Small‐Scale Farming: A Non‐Parametric Analysis of the (Inverse) Relationship Between Farm Productivity and Farm Size in Burundi." *Journsal of Agricultural Economics* 64.2 (2013): 319-342.

[41] Wang, Jianying, et al. "Is small still beautiful? A comparative study of rice farm size and productivity in China and India." *China Agricultural Economic Review* 7.3 (2015): 484-509.

Appendix

Table2 Parametric results from Semi-parametric estimation

	Semi-parametric Reg.	
VARIABLES	(1) yield	(2) profit
Disaster	-166.2***	-241.5***
	(-13.45)	(-13.89)
Property	-20.18	-30.14
	(-1.44)	(-1.54)
Land slope	100.4***	106.9***
	(7.01)	(5.32)
Land quality (basement=1,Good)		
Medium	-60.30***	-38.32**
	(-5.06)	(-2.36)
Bad	-105.4***	-106.4***
	(-6.09)	(-3.96)
Irrigation	71.57***	64.19***
	(5.12)	(3.16)
Soil type	-0.89	-4.31
	(-0.07)	(-0.23)
Distance	-0.55	-0.943
	(-0.26)	(-0.26)
Plots number	-0.167	0.227
	(-0.32)	(0.51)
Education	4.837***	2.291
	(2.58)	(0.81)
Gender	-32.03*	-37.49
	(-1.89)	(-1.58)
Age	1.189**	1.075
	(2.07)	(1.31)
Leader	50.44***	67.62***
	(4.17)	(4.04)
Family size	0.954	-0.257
	(0.32)	(-0.06)
Labor rate	-12.829	-73.16**

	(-0.50)	(-2.11)
Year	135.0***	-48.81*
	(6.66)	(-1.80)
Province Dummy	controlled	controlled
Observations	1,845	1,845
Adj R-squared	0.3605	0.3429
Root MSE	218.8940	313.8181

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table3 Results of Hardle and Mammen's test

	(1) Yield	(2) Profit
Standardized Test statistic T	2.892	3.510
Critical value (95%)	1.96	1.96
Approximate P-value	0.00	0.00