Technical Efficiency and Optimal Farm Size in the Tajik's Cotton Sector

Tsimpo Nkengne, Clarence

University of Montreal

10 March 2010

Technical Efficiency and Optimal Farm Size in the Tajik's Cotton Sector[†]

Clarence Tsimpo Nkengne[ŧ]

March 2010

Abstract: The main objective of this paper is to estimate the technical efficiency of cotton farms in Tajikistan using a stochastic frontier production function, and to derive the optimal farm size. Currently, Tajikistan is reforming its cotton sector. This reform consists essentially of switching from a communist system with large state owned farms to a private system. This brings the question of what the optimal size of the new private farms should be. The study involved collection and analysis of data on 205 cotton farms from the Sughd province where cotton production is concentrated. The analysis suggests that an inverse relationship between productivity and farm size does not hold. The relationship between farm size and technical efficiency is more complex than what is normally believed.

Key Words: Agriculture, Farm size, Stochastic frontier model, Tajikistan

JEL Classification:

[ŧ]Clarence Tsimpo Nkengne: Department of Economics, Université de Montréal, C.P.6128, succ. Centre-Ville Montréal, Québec, Canada H3C 3J7. Email: clarence.nkengne.tsimpo@umontreal.ca.

[†]I am thankful to CECI for the opportunity that I was given to participate on the project. At various stages of this paper, I benefited from discussions with Andriana Bellou, Quentin Wodon, Etienne Lamy, Prospere Backiny, George Joseph, Jean Pascal Nganou, Kofi Nouve, ...

I. Introduction

Agriculture has been a focus in the literature on development for a long time, in part because in developing countries the economy is often driven by agriculture, in terms of employment, share in GDP, public revenue, and export. Given the recent worldwide food crisis, the role of agriculture in the process of economic development has received renewed attention in the ongoing debate on how to improve productivity and efficiency. This is the case in Tajikistan where the Government in coordination with donors is implementing reforms in order to enhance agricultural performance, especially in the cotton sector. One of the main issues in the reform is related to farm size. It has been well documented in the economics literature that the size of the farm matters for efficiency and productivity, but there is no consensus on what the optimal size should be.

The purpose of this paper is to estimate empirically the technical efficiency (Te) of cotton production in Tajikistan, to derive the relationship between technical efficiency and farm size and to estimate the optimal farm size for cotton production. This should be of interest to policymakers and the ongoing land reform debate. The theoretical discussion of efficiency started with Farrell (1957). Recent work includes Kompas and Nhu Che (2006) who use the stochastic frontier to understand the dynamic of efficiency of milk production in Australia following a suppression of governmental subsidies; Igbekele and Al. (2006) use the same approach to compare the efficiency among rural and urban producers in Nigeria.

A popular stylized fact in development economics is that there is a strong inverse relationship between farm size and land productivity (Sen, 1962). The inverse relationship is typically explained by the difference in factor endowments between small and large farms: by using family labor smaller farms face lower labor transaction costs than larger farms (Shenggen Fan and Chan-Kang, 2005). As a result, smaller farms have higher labor/land ratios and can achieve higher yields per hectare. The inverse relationship has important implication for land reform policy, as it is argued that any type of land reform that reduces inequality in landholdings will likely have positive effects on productivity. The question is whether this is the case also in Tajikistan's cotton sector.

Early studies on the question of productivity and farm size include Alexander Chayanov (**1920**) who stated that the size of the farm is positively correlated to the size of the household. Sen (1962) observed that small farmers were more productive per unit of land than large farmers. However, with the advent of the Green Revolution, research has shown that the relation diminishes or is even reversed as agriculture become more capital intensive (Shenggen Fan and Chan-Kang, 2005). One of reasons why the inverse relationship broke down relates to preferential access by large farms to institutions and services that help lower inefficiency such as rural electrification, technical assistance, access to markets as well as more use of intensive technologies and inputs that raise productivity (Helfand and Levine, 2004). Empirical evidence suggested that there has been different dynamics across countries over time (see Robert Eastwood, Michael Lipton and Andrew Newell, 2004).

This study involved collection and analysis of data on 205 cotton producers from Sughd province where cotton production is concentrated. Based on estimates of a stochastic frontier model (Battese and Coelli, 1992), efficiency scores are computed for each farm. Then I look at the distribution of technical efficiency across farm size deciles and derive a plot using kernel regression of the technical efficiency as a function of farm size.

The results suggest that on an average, farmers in Tajikistan tend to realize about 70 percent of their technical capability. The relationship between farm size and technical efficiency is however more complex than the inverse relation suggested by most of the literature. The optimal cotton farm size in Tajikistan is around 53 and 56 ha. The findings of this paper complement the literature on optimal farm size in general and in particular the more recent research on farm size and productivity by Helfand and Levine (2004) and Shenggen Fan and Chan-Kang (2005). The findings of this paper suggest that could be both an inverse and a reverse U shape relationship between farm size and efficiency.

The paper is organized as follows. Section II presents the model used for the stochastic production frontier. Section III describes the data and analyzes the empirical results with an emphasis on farm size. The last section provides concluding remarks.

II. Analytical Framework

For this study, the stochastic frontier production function is used to estimate the technical efficiency for the sample farmers. Efficiency of a production system or unit means a comparison between observed and optimal values of its output and inputs. The comparison can take the form of the ratio of observed to maximum potential output obtainable from the given inputs. In this comparison, the optimum is defined in terms of production possibilities, and efficiency is technical. A farm is said to be technically inefficient if too little output is being produced from a given bundle of inputs. Hence, inefficiency involves excessive usage of all inputs.

Following pioneering but independent works by Aigner and al. (1977), Battese and Corra (1977) and Meeusen and van den Broeck (1977), it is now feasible to estimate frontier production functions relativelly easily. The idea of a frontier function can be illustrated with a deckam farm using k inputs $(X_1, X_2, ..., X_k)$ to produce output Y. Efficient transformation of inputs into output is characterized by the production function $f(X_i)$ which shows the maximum output obtainable from various input vectors. The stochastic frontier production function assumes the presence of technical inefficiency of production. Hence the function is defined by,

$$Y_i = f(X_i, \beta) \exp(V_i - U_i) \qquad i = 1, 2, ..., N. \tag{II.1}$$

Where N is the number of farmers in the sample, Y_i is the production by hectare of the ith farmer, X_i is the input quantity by hectare, β's are production coefficients. The error component U_i is assumed to be distributed independently of V_i, and to satisfy $U_i \geq 0$. V_i is a random error, which is associated with measurement error and random factors not under the control of the farmer such as luck, climate, topography, strikes, and machine performance. The inefficiency measure, U_i is itself affected by other variables under the farmer's control, such as knowledge and effort.

When the farm is fully technically efficient, U_i takes the value of 0 and when the farm is inefficient U_i takes a value greater than 0. The magnitude of U_i specifies the "efficiency gap", that is how far a farm's given output is from its potential output.

Choosing an appropriate distributional form for the U_i's is a difficult task because, in doing so, the researcher is assuming to know quite a lot about the unknown phenomenon under investigation. Greene (1993) presents several explicit forms that refer to different assumptions about the distribution of the inefficiency term. Most commonly used one-sided distributions are the exponential, the half-normal and the truncated normal distributions. The most frequently used form is to assume that U_i is independently and identically distributed and truncated at zero of the normal distribution with mean μ and variance σ_u^2.

Two common forms of production functions are used to estimate technical efficiency using the stochastic frontier production function, namely Cobb-Douglas and general translog functional forms. The Cobb-Douglas frontier model describing the production of farmers is given by

$$Y_i = \beta_0 + \sum_{j=1}^{5} \beta_j X_{ji} + V_i - U_i. \tag{II.2}$$

Where the subscript i represents the ith farmer; Y_i represents the logarithm of the physical output of the ith farmer per hectare; and in our case X_1 represents the logarithm of the quantity of seeds used per hectare; X_2 is the logarithm of family labor and hired labor per hectare; X_3 represents the logarithm of the quantity of fertilizers used per hectare; X_4 represents the logarithm of the quantity of treatment products used per hectare; X_5 represents the logarithm of the value of the capital (tractors and others equipments) owned by the farmer.

The V_i's and U_i's are as defined earlier. U_i is the non-negative truncation (at zero) of the normal distribution[1] with mean, μ_i, and variance, σ^2, where μ_i is defined by,

$$\mu_i = \delta_0 + \sum_{m=1}^{3} \delta_m Z_{mi} = Z_i \delta. \tag{II.3}$$

[1]Guarantees inefficiency to be positive only.

Where Z_1, Z_2 and Z_3 represent the proportion of cotton area in total land, seedbed quality and managerial knowledge respectively,[2] which are assumed to influence the technical efficiency of farmers. These three variables are included in the model as determinants of technical inefficiency to indicate possible effects of farm and farmers' characteristics on the efficiency of production.[3] The efficiency score of the ith farmer, given the specifications of the model, is defined by $Te_i = exp(-U_i)$.[4]

III. Empirical Results

Data for the study were collected from an agro-economic survey conducted in 2006 with one observation per farm, so that it is not possible to use panel data for estimating the stochastic production frontier. The data were collected by the Canadian Center For International Studies and Cooperation (CECI[5]). The study area covered all the districts of Sughd province. Except for a few collective farms which are still under the control of the State, the sample is representative of the population of farms in the province. The selection of these farms was done in two steps. The first step involved field work (administrative data collection and verification if needed) in order to create a sampling frame. At the second step, the sample was draw from the sampling frame. The probability of selection was the farm size. All surveyed farms produced cotton. A total of 205 farms, distributed over the various districts, were interviewed. The survey obtained data on land use, agricultural production, irrigation practices, management, input levels, labor, processing and marketing, and use of credit.

Table 1 presents summary statistics for the variables used in this analysis. They include the sample mean values and the standard deviation, together with the minimum and maximum values of each of the variables. The typical farm cultivates 30 hectares of cotton.

[2] Managerial knowledge is an indicator based on adequacy of the planning process; effectiveness of accounting system; availability and access to technical support services and adequacy of input supply services. Seedbed quality is an indicator of the quality of the seedbed after soil preparation.

[3] For sensitivity analysis, a parallel model is estimated with no explanatory variable for U_i.

[4] See appendix for details on the estimation process.

[5] Centre d'Etude et de Coopération Internationale.

But there are lots of disparities as the size range from 1 to 351 hectares. This confirms the assumption that the farm size is a key issue for the ongoing reform. The typical farm production of unprocessed cotton is estimated at 59.6 tons. This results in low yields (1.94 ton per ha), and hence low incomes for cotton farmer. There is room for improvement a World Bank (2004) report suggest that cotton yield could reach 3 tons per ha as was the case in the early 1990.[6]

Table 1. Summary statistics for the characteristics of farms

The maximum likelihood estimates of the model are presented in Table 2. For sensitivity purposes, I have estimated six models. The models differ on two grounds: (i) the assumption regarding the distribution of U_i's; (ii) and the presence or not of explanatory variables for U_i. The six models give highly correlated results. I will focus then on the first model. The coefficients of the input variables in the Cobb-Douglas production function are the elasticities of mean output with respect to the different inputs used. All elasticities are positive as expected. The elasticity of frontier (best practice) production with respect to seeds is estimated to be 0.4096. Thus if the quantity of seeds per hectare were to be increased by 1 percent, cotton yields would increase by 0.4096 percent. The elasticity of human labor is estimated at 0.0538. The elasticity of output with respect to fertilizers, treatment products, and machinery are lower but also highly statistically significant. The return to scale parameter for the Cobb- Douglas production frontier is estimated as the sum of the elasticities of the five inputs. This suggests that cotton cultivation in Tajikistan experiences decreasing returns to scale, since the sum of the input elasticities is lower than one.

The estimated coefficients of the explanatory variables for technical efficiency are of particular interest. Each of the variables in the efficiency model has a negative sign implying that an increase in the value of these variables would increase technical efficiency.

The negative estimate for the proportion of the cultivated area allocated to cotton

[6] This yield is far from world best. The average yield for Australia, the world top performing country, is 4.4 tons/ha and 3.9 tons/ha for China (in 2007 and according to FAOSTAT website).

implies that farmers with a greater proportion of land dedicated to cotton tend to be less inefficient. The negative coefficient for seedbed quality suggests that inefficiency tends to decline with seedbed quality. The negative coefficient for manager knowledge indicates that inefficiency decreases with managerial abilities, but this coefficient is not statistically significant. The estimate for the variance parameters σ_u, σ_v,and γ(close to one), indicate that the inefficiency effects are likely to be highly significant overall.

Table 2. Maximum likelihood estimates of parameters of the frontier models

The predicted mean technical efficiencies for cotton farms were estimated to be between 0.670 and 0.738 depending of the model (Table 3). Thus farmers in Tajikistan tend to realize about 70 percent of their potential production. To elaborate on the optimal farm size, I will look at the distribution of efficiency across area deciles.

On table 3, mean efficiency estimates are provided for the six models considered, and confidence interval are provided in table 4 for the preferred model.

A combination of factors is likely to drive the efficiency of farms. Managers of the most efficient farms tend to have a higher knowledge index. These farms used few laborers per hectare. This is a sign of intensive mechanization. Also, the optimal farms present an appropriate balance in seeds, fertilizers and treatment products. The inverse relationship does not hold here. This can be explained by a high entrance cost to mechanization. When the farm is too small, the inverse relation holds, but as the size increase, given the mechanization, the efficiency/yield increase, until one reach a maximum around 56 ha. Then the inverse relation holds again.

Table 3. Mean efficiency by area deciles

Table 4. Dispersion of confidence interval by farm size for model 1

Table 5. Distribution of keys variables across area deciles

IV. Conclusion

The purpose of this paper was to analyze the efficiency of the Tajik cotton sector and to come out with the optimal farm size. The analysis was based on the data from a comprehensive survey on 205 farms, and the use of the stochastic frontier production function. Two broad messages emerge from our analysis. First, although the reform is undergoing since a long time, the production of cotton in Tajikistan is still facing urge difficulties with as consequence, a poor productivity and efficiency. This is a confirmation that the reforms were needed, and that at lot is still to be done to put down the Bottleneck facing by the sector. Second, and most importantly, I find that the relationship between farm size and technical efficiency is more complex than what is normally believed. The optimal farm size is around 53 and 56 ha. This is a key result that must guide the debate on the ongoing land reform in Tajikistan and to hold the downsizing of the farm size to and appropriate level. An important improvement would be to take into account the impact of farm on climate change and see how this may affect the result.

V. Appendix: Estimating the parameters

The likelihood function is expressed in terms of the variance parameters, $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\gamma = \sigma_u^2 / [\sigma_u^2 + \sigma_v^2] \in [0, 1]$ (or $\lambda = \sigma_u/\sigma_v$). Note that when $\gamma = 0$, deviations from frontier are due entirely to noise. When $\gamma = 1$, deviations from frontier are due entirely to inefficiency. This parameterization has advantage that we can search for values of γ over $[0, 1]$ as start value for iterative maximization step. The distribution function of the sum of a symmetric normal random variable and a truncated normal random variable was first derived by M. A. Weinstein (1964). The derivation of the density function of $\varepsilon_i = (V_i - U_i)$ is straightforward:

$$f_\varepsilon(\varepsilon_i) = \phi\left(\frac{\varepsilon_i}{\sigma}\right)\left[1 - \Phi\left(\frac{\varepsilon_i \lambda}{\sigma}\right)\right], \ldots - \infty \leq \varepsilon_i \leq +\infty. \tag{V.1}$$

Where ϕ and Φ are the standard normal density and distribution functions, respectively.

The result is Following Battese and Coelli (1992), we can detail the log-likelihood under this parameterization as follow:

$$\log L = -\frac{N}{2}\log(\pi/2) - \frac{N}{2}\log(\sigma^2) + \sum_{i=1}^{N}\log\left[1 - \Phi\left(\zeta_i\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(Y_i - X_i\beta\right)^2 \qquad \text{(V.2)}$$

where

$$\zeta_i = \frac{(Y_i - X_i\beta)}{\sigma}\sqrt{\frac{\gamma}{1-\gamma}}$$

Where $\Phi(.)$ is the distribution function of a standard normal random variable. The two most commonly used methods of estimating the parameters of a stochastic frontier are maximum likelihood estimation (MLE) and corrected ordinary least squares (COLS).[7] The method of maximum likelihood is proposed for simultaneous estimation of the parameters of the stochastic frontier and the model for the technical inefficiency effects.

In the case of the one-parameter exponential distribution for U_i's,

$$f(U_i) = \frac{1}{\theta}\exp\left(U_i/\theta\right), \ldots U_i \geq 0.$$

Where $\theta \geq$ is the mean of $-U_i$ (The variance is θ^2). The density function of $\varepsilon_i = (V_i - U_i)$ is given by:

$$f_\varepsilon(\varepsilon_i) = \frac{1}{\theta}\left[1 - \Phi\left(\frac{\varepsilon_i}{\sigma_v} + \frac{\sigma_v}{\theta}\right)\right]\exp\left[\frac{\varepsilon_i}{\theta} + \frac{\sigma_v^2}{2\theta^2}\right], \ldots -\infty \leq \varepsilon_i \leq +\infty. \qquad \text{(V.3)}$$

The log-likelihood function for the model under the exponential parameterization follows.[8]

The useful parameters required to estimate residuals ε_i can be easily obtained by OLS, or by maximizing the log-likelihood. However, the problem of decomposing ε_i into its

[7] The method uses the moments of the OLS residuals to calculate an estimate of γ (or λ) and then uses this value to adjust the OLS estimates of β_0 and a σ^2 (Coelli, 1995).

[8] See Aigner and al. (1977) for more details.

components V_i and U_i still remain. This issue is solved by considering the conditional distribution of U_i given ε_i.[9]

The estimation of the maximum likelihood uses a three-step estimation procedure. The first step involves calculation of OLS estimates of β. These estimates are unbiased estimators of the parameters in equation (II.2), with the exception of the intercept, β_0, and σ^2 (Aigner et al, 1977). In the second step, a grid search of γ is conducted. The likelihood function is evaluated for a number of values of $\gamma = \sigma_u^2/(\sigma_u^2+\sigma_v^2)$ between zero and one. Any other parameters (μ or δ's) are set to zero in this grid search. The β_0, and σ^2 parameters adjusted according to the corrected ordinary least squares formula presented in Coelli (1995). The final step uses the best estimates (that is, those corresponding to the largest log-likelihood value) from the second step as starting values in a Davidon-Fletcher-Powell (DFP) maximization routine which produces the final maximum likelihood estimates.

Battese & Coelli (1993) show that for the ith farm, the technical efficiency is predicted using the conditional expectation,

$$\begin{aligned} Te_i &= \exp(-U_i|\varepsilon_i) \\ &= \exp\left(-\mu_* + \frac{1}{2}\sigma_*^2\left(\frac{\phi\left[(\mu_*/\sigma_*) - \sigma_*\right]}{\mu_*/\sigma_*}\right)\right) \end{aligned}$$

Where

$$\mu_* = (1-\gamma)\, Z_i\delta - \gamma\varepsilon_i, \ldots \sigma_*^2 = \gamma\,(1-\gamma)\,\sigma_v^2, \ldots \varepsilon_i = V_i - U_i$$

References

1. Aigner, Lovell and Schmidt. (1977). "Formulation and estimation of stochastic frontier production models" Journal of Econometrics, 6, pp. 21–37.

[9]See Bera and Sharma (1999) for details on the conditional distribution of U_i given ε_i. $f(U_i|\varepsilon_i)$ is also useful if one need to compute the confidence intervals for U_i.

2. Ana R. Rios and Gerald E. Shively (2005), "Farm size and nonparametric efficiency measurements for coffee farms in Vietnam",Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Providence, Rhode Island, July 24-27, 2005.

3. Anil K. Bera and Subhash C. Sharma (1999), "Estimating production uncertainty in stochastic production frontier", Journal of Productivity Analysis, Volume 12, Number 3 / novembre 1999.

4. Ardilly Pascal (2006). "Les techniques de sondage" Editions Technip.

5. Alexander Chayanov (1919), "The Theory of Peasant Co-operatives"

6. CECI (2007), "Benchmarking of DF based on efficiency and profitability"

7. Coelli, T.J. (1995), "Estimators and Hypothesis Tests for a Stochastic: A Monte Carlo Analysis", Journal of Productivity Analysis, 6, 247-268.

8. Dussaix A. M. & J. M. Grosbras. (1993). "Les sondages : principes et méthodes" Presses Universitaires de France.

9. G.E. Battese and T.J. Coelli. (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India" The Journal of Productivity Analysis, 3, 153-169.

10. George E. Battese. (1992). "Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics" Agricultural Economics, 7 (1992) 185-208.

11. Gourieroux Christian. (1981). "Théorie des Sondages" , 270 p., Economica.

12. Igbekele A. Ajibefun, Adebiyi G. Daramola &Abiodun O. Falusi. (2006). "Technical Efficiency of Small Scale Farmers: an Application of the Stochastic Frontier Production Function to Rural and Urban Farmers in Ondo State, Nigeria" International Economic Journal, Vol. 20, No. 1, 87–107.

13. J.P. Boussemart, J.P. Butault and E. Matvejef (2006), "Economies of scale and optimal farm size in the Estonian dairy sector", 96 th EAAE-seminar January 2006, Taenikon, Switzerland."Causes and Impacts of Agricultural Structure".

14. Marchand, Sébastien (2009). “technical efficiency, farm size and tropical deforestation in the Brazilian amazonian Forest” MPRA, paper No 13648 .

15. Meeusen, W.&van den Broeck, J. (1977). “Efficiency estimation from Cobb–Douglas production functions with composed error” International Economic Review, 18, pp. 435–444.

16. Pranab Bardhan and Dilip Mookherjee (2007), "Land Reform and Farm Productivity in West Bengal", Working paper

17. Robert Eastwood, Michael Lipton and Andrew Newell (2004), "Farm Size", Paper prepared for Volume III of the Handbook of Agricultural Economics.

18. Tom Kompas and Tuong Nhu Che. (2006). “Technology choice and efficiency on Australian dairy farms” The Australian Journal of Agricultural and Resource Economics, 50, pp. 65–83.

19. Sen, A. K.(1962), “An Aspect of Indian Agriculture”, Economic Weekly 14, No 4-6, pp 243-246.

20. Steven M Helfand and Edward S. Levine (2004), “Farm size and the determinants of productive efficiency in the Brazilian Center-West” Agricultural Economics, 31, pp. 241-249.

21. Shenggen Fan and Connie Chan-Kang (2005), “Is small beautiful? Farm size, productivity, and poverty in Asian agriculture” Agricultural Economics, International Association of Agricultural Economists, vol. 32(s1), pages 135-146, 01.

22. World Bank (2004), “PSIA of Tajik cotton sector”.

Table 1. Summary statistics for the characteristics of farms

Variable		Obs	Mean	Std, Deviation	Min	Max
Cotton area	Ha	205	30.8	42.4	1.0	351.0
Production (unprocessed cotton)	Tons	205	59.6	72.0	1.3	600.0
Commercial seeds	Kg	205	3 695.7	4 572.1	70.0	33 000.0
Laborers + Hired laborers	Number of persons	205	68.0	114.0	1.0	1 000.0
Laborers	Number of persons	205	58.8	109.0	1.0	900.0
Hired laborers	Number of persons	205	9.2	18.0	0	100
Fertilizers	Kg	205	40 364.3	81 820.4	352.0	836 000.0
Treatment products	Liter	205	83.8	293.6	0.0	2 200.0
Machinery	Som	205	13 435.0	21 650.3	0.0	241 638.1
% not using treatment products	%	205	0.53	0.5	0.0	1.0
Proportion of cotton area	%	205	65.6	20.9	5.9	100.0
Seedbed quality	%	205	81.0	26.0	0.0	100.0
Manager knowledge	%	205	35.6	48.0	0.0	100.0

Note: The table reports the basics statistics of the sample farms.

Table 2. Maximum likelihood estimates of parameters of the frontier models

	Exponential		**Half-normal**		**Truncated-normal**	
Cotton yield	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Commercial seeds per ha	0.4096***	0.4434***	0.4798***	0.4817***	0.4095***	0.4252***
	[0.0906]	[0.0913]	[0.0804]	[0.0827]	[0.0858]	[0.0920]
Labor + Hired labor per ha	0.0538**	0.0487*	0.0644**	0.0610**		
	[0.0272]	[0.0276]	[0.0271]	[0.0276]		
Fertilizers per ha	0.0634***	0.0644***	0.0675***	0.0650***	0.0814***	0.0757***
	[0.0209]	[0.0213]	[0.0215]	[0.0219]	[0.0202]	[0.0206]
Treatment products per ha	0.0347**	0.0415**	0.0369**	0.0444**		
	[0.0153]	[0.0163]	[0.0167]	[0.0184]		
Machinery per ha	0.0611***	0.0578***	0.0653***	0.0612***	0.0693***	0.0643***
	[0.0194]	[0.0193]	[0.0193]	[0.0191]	[0.0189]	[0.0190]
Constant	-1.8261***	-1.9566***	-2.1352***	-2.0836***	-1.9108***	-1.9336***
	[0.4737]	[0.4742]	[0.4266]	[0.4334]	[0.4517]	[0.4787]
lnsig2v	-3.6891***	-3.7156***	-4.0440***	-4.1511***		
Constant	[0.2770]	[0.2839]	[0.3366]	[0.3478]		
lnsig2u					**mu**	
Proportion of cotton area	-0.0204***		-0.0149***		-0.0235*	
	[0.0066]		[0.0045]		[0.0136]	
Seedbed quality	-0.0140**		-0.0094**		-0,0121	
	[0.0060]		[0.0042]		[0.0079]	
Manager knowledge proxy	-0,0048		-0,0036		-0,0031	
	[0.0033]		[0.0023]		[0.0036]	
Constant	0,4727	-1.9816***	0,6422	-1.0480***	1.5827***	-8,5790
	[0.6415]	[0.2001]	[0.4379]	[0.1320]	[0.5202]	[27.8767]
ilgtgamma						
Constant					3.3258***	4.9258*
					[0.6102]	[2.8275]
lnsigma2						
Constant					-0,4202	1,2790
					[0.5973]	[2.9092]
Observations	205	205	205	205	205	205
Sigma_u	-	0.371	-	0.592	0.796	1.889
Sigma_v	0.158	0.156	0.132	0.125	0.151	0.161
Log Likelihood	-63.23	-72.94	-64.92	-75.91	-66.00	-77.46

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1
Note: The table reports the coefficients of the stochastic frontier production function. A total of six models where estimated. The estimated efficiency's are highly correlated. Model 1 is our preferred model. The elasticity of frontier (best practice) production with respect to seeds is estimated to be 0.4096. This indicated that, if the quantity of seeds per hectare were to be increased by 1 percent, then cotton yield were estimated to increase by 0.4096 percent. Further, the elasticity of human labor is estimated to be between 0.0538. Base on elasticity, fertilizers are the second most important input. The elasticity of output in respect of fertilizer, treatment products and machinery are as low as the one of labor, and also highly significant. The return to scale parameter for the Cobb- Douglas production frontier is estimated by the sum of the elasticity's of the five variables. It is found that the cotton cultivation in Tajikistan experienced decreasing returns to scale, as the sum of input elasticity's was lower than one. Note that all variables in the production function are per hectare. The estimate for the variance parameters σ_u , σ_v , and γ(close to one), indicates that the inefficiency effects are likely to be highly significant in the analysis of the value of output of the farmers.

Table 3. Mean efficiency by area deciles

Cotton area Deciles	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Lower	0.731	0.730	0.679	0.673	0.714	0.734
2	0.782	0.772	0.724	0.711	0.765	0.775
3	0.702	0.694	0.640	0.628	0.664	0.682
4	0.770	0.766	0.710	0.700	0.751	0.769
5	0.575	0.564	0.516	0.506	0.532	0.546
6	0.800	0.794	0.757	0.747	0.767	0.776
7	0.722	0.706	0.651	0.635	0.696	0.706
8	0.813	0.796	0.760	0.747	0.787	0.788
9	0.833	0.817	0.773	0.760	0.794	0.799
Higher	0.658	0.654	0.605	0.596	0.631	0.646
All DF	0.738	0.728	0.681	0.670	0.710	0.722

Note: The table reports the efficiency score from the six models across area deciles. Technical efficiency varied from 0.670 to 0.738. The 9th decile is always the most efficient.

Table 4. Dispersion of confidence interval by farm size for model 1

	Model 1			
Cotton area Deciles	Mean Efficiency	Lower CI	Upper CI	Range
Lower	0.731	0.653	0.809	0.157
2	0.782	0.723	0.841	0.118
3	0.702	0.625	0.778	0.154
4	0.770	0.710	0.831	0.121
5	0.575	0.480	0.670	0.190
6	0.800	0.716	0.884	0.168
7	0.722	0.651	0.792	0.141
8	0.813	0.738	0.888	0.149
9	0.833	0.786	0.880	0.095
Higher	0.658	0.555	0.762	0.207
All DF	0.738	0.712	0.764	0.052

Note: Note: The table reports the efficiency score across area deciles. The width is however wider for the dispersion of confidence intervals on farm size basis, where the range is between 0.095 and 0.207. The highest width of intervals (0.207) is with the group of farms with the highest size (120 hectares on average for this group), while the least width (0.095) is among the group just behind the highest decile (for this group the average size is 56 hectares). This group is also the one with the higher efficiency score (0.833). The implication of this result is that the farms optimal size may be around 56 hectares.

Table 5. Distribution of keys variables across area deciles

	Area Deciles										
	Lower	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Higher**	**All DF**
Efficiency	0.731	0.782	0.702	0.770	0.575	0.800	0.722	0.813	0.833	0.658	0.738
Cotton area (Ha)	2.38	4.37	6.47	8.95	14.24	21.48	31.83	45.55	56.23	119.50	30.77
Yield (Tons/Ha)	2.01	2.14	1.78	2.10	1.46	2.10	2.00	2.38	2.42	1.53	1.94
Commercial seeds (Kg/Ha)	100.32	106.52	111.27	108.30	133.93	104.24	134.36	129.99	130.97	110.29	120.12
DF Laborers + Hired Laborers(Persons/Ha)	5.72	4.34	2.02	4.05	1.79	1.72	2.48	1.75	1.62	2.45	2.21
DF Laborers (Persons/Ha)	2.28	2.25	1.37	2.01	1.34	1.32	2.27	1.58	1.49	2.32	1.91
Hired Laborers (Persons/Ha)	3.44	2.09	0.64	2.04	0.45	0.40	0.22	0.17	0.13	0.13	0.30
Fertilizers (Kg/Ha)	2804.16	1920.90	1907.84	2683.60	2196.23	755.14	2240.46	1214.12	1453.19	815.51	1311.97
Treatment products (Liter/Ha)	0.58	10.33	3.84	1.98	0.53	0.61	5.19	5.83	3.70	0.70	2.72
Capital (Som/Ha)*	616.06	441.70	519.38	427.26	462.02	426.43	512.70	443.82	496.23	375.68	436.68
Proportion of cotton area (%)	58.86	69.06	70.54	62.50	58.35	64.53	67.14	77.68	72.68	54.79	65.58
Seedbed quality (%)	79.37	79.71	62.96	73.33	78.79	84.21	85.71	92.06	96.67	75.00	80.98
Manager Knowledge (%)	9.52	13.04	16.67	30.00	36.36	47.37	47.62	61.90	50.00	45.00	35.61

Note: The table reports the mean of the keys variables by area deciles. The most efficient decile is the 9th, with a yield of 2.42 tons/hectare.