International Journal of Agricultural Management and Development (IJAMAD)
Available online on: www.ijamad.iaurasht.ac.ir
ISSN: 2159-5852 (Print)
ISSN:2159-5860 (Online)

Evaluation of Palm Groves Technical Efficiency Using Bootstrap Data Envelopment Analysis: A Case Study of Roodkhanehbar Area, Iran

Gholamreza Zamanian [1] and Mostafa Khajeh Hassani [2*]

Received: 26 August 2015,
Accepted: 08 Octber 2015

Abstract

Roodkhnehbar area, having approximately 111 thousands of Keriteh palm trees, is one of the most important areas of date production in the Rudan County[1] and the source of peoples' income in this area, directly or indirectly. As a result, its production efficiency has a critical importance to the orchardists in this region. This study aims to evaluate technical efficiency of palm groves in this area using input-oriented bootstrap data envelopment analysis and sampling 50 palm groves of Keriteh date producers of Roodkhanehbar area in 2013. The results suggested that 64% of date producers operate with less than 50% efficiency and only 14% of them operate efficiently. The study, then, carries the implication that it is recommended to train the orchardists, providing a chance for successful orchardists to share their experiences with others in an attempt to optimize allocation of inputs.

Keywords:
bootstrap, Data Envelopment Analysis, Hormozgan Province, Keriteh date, Roodkhanehbar, technical efficiency

[1] Assistant Professor, Economics, University of Sistan and Baluchestan, Zahedan, Iran
[2] PhD Candidate, Agricultural Economics, University of Sistan and Baluchestan, Zahedan, Iran
* Corresponding author's email: zamanian@eco.usb.ac.ir

[1] Rudan County is a county in Hormozgan Province in Iran

INTRODUCTION

One of the most important economic goals and policies implemented in Iran is to decrease dependency on oil revenues and increase non-oil products revenues. Considering oil market depression in recent years and harmful effects which single-product export has on the country, the importance of this policy becomes further egregious. Taking this into account and because of high potentials in agricultural sector, the country policy makers and economists should consider these facts in their planning. Agricultural sector of Iran includes a wide range of activities and the most important of them is date production, which as a potential substitute, can provide more work opportunities and thereby support the development of the country. As shown below (Figure 1), producing 1.061 million tons of dates in 2012, Iran had the world's second rank in date production after Egypt. In addition to meeting domestic demands andexporting 112030 tons of dates in 2011 Iran is ranked as the third country after Iraq and Pakistan,.

Despite Iran's rank in production and export of dates, this country was ranked twelfth among 37 producing countries in 2012, and has had not great performance per unit area compared to other competitors such as Egypt, China, Sudan, Colombia, and America, because of inappropriate and traditional production methods in producing chain in nation groves (FAO, 2012). For this reason, to maintain the position and increase comparative advantages, and subsequently, increase the competition power in foreign markets, it is always necessary that the performance should be improved by studying efficiency and productivity indexes of groves of different regions of the country.

Hormozgan Province is one of the most important producers of dates in the country. There are over 100 varieties of dates in the country, 80 of which are planted in Hormozgan palm gardens. The most important and favorable varieties include Mazafati, Khanizi, Shahani, Piaram, Khasoei, Keriteh, Mordasang, Zarak,and red Kolk. This province is located in fourth and seventh ranks, respectively in terms of palm cultivation and production. It takes five months to harvest dates in this and some other provinces. Roodkhanehbar area, located in Rudan County, is one of the palm growing areas where this activity has so much influence on the quality of residents' lives. This area has the biggest area under Keriteh date cultivation with about 100 groves and over one hectare and 111 thousands palm trees. Despite being very nutritious, Keriteh variety is not considered among other expensive varieties such as Piaram and Mazafati and is produced as human, livestock and industrial food supply in foreign markets of China, India, and Afghanistan. The majority of the people in this region are laborers at palm groves, and any improvement in the allocation of resources which leads to the reduction of production costs, will consequently increaselabor profit, which will considerably accelerate the development of the region. What is of great importance is evaluation of technical efficiency of the region palm grovesas well as identification of inefficient units; theselead to improvement in resource allocation. Given the perceived necessities, the present study aimed to evaluate the technical efficiency of Roodkhanehbar regionKeriteh palm groves in 2013 using bootstrap input oriented data envelopment analysis. Despite the fact that extensive studies have already been conducted

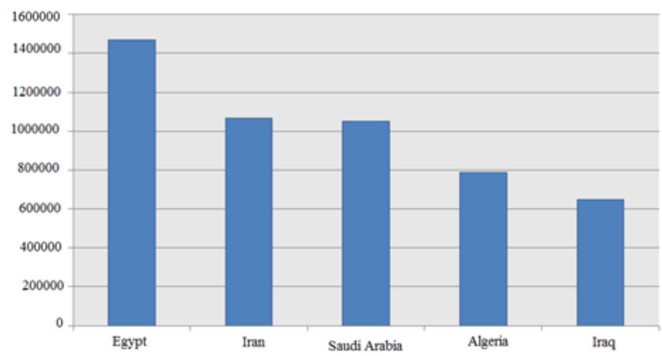

Figure 1. The first five date producers of the world

in the field of evaluating agricultural sector activities efficiency inside and outside the country, a few of them have used bootstrap data envelopment analysis, the most important of them include (Balcombe et al., 2008; Brümmer, 2001; Dong & Featherstone, 2006; Gocht & Balcombe, 2006; Odeck, 2009). A theoretical foundations and methodology of study;brief description of the data used and the results of model estimation, are presented here.

MATERIALS AND METHODS

Theoretical foundations

In 1957 Farrell calculated efficiency of American agricultural sector based on Economic theories using nonparametric methods. Citing five principles, he constructed a collection called production possibility and considered a part of its frontier as an estimation of production function. Every Decision Making Unit (DMU), which is placed on this frontier, is efficient and inefficient otherwise (Farrell, 1957). Because of scientific problems in measuring and limitations raised in the Farrell method, it did not find much practical application and remained silent for years until 1978. To remedy this problem, Charnes, Cooper, and Rhodes (CCR) introduced the method of Data Envelopment Analysis (DEA) by universalizing the Farrell method such that it includes the characteristic of production process with several factors and outputs. In this method, it is not required to follow a certain default order so as to estimate the production function, and efficiency of a firm (decision making unit) is measured relative to the efficiency of other firms (Cooper et al., 2006). Unfortunately, in actual situation the maximum amount of output produced from a particular input cannot be calculated by using a sample, because the studied sample is a sample of an unknown population, and efficient frontier of the population is unknown. As shown in the following figures, the efficiency of unit A in Figure 2 significantly changes compared to Figure 3 (Bahadori et al., 2013).

This efficiency change is a consequence of nonparametric nature of DEA model. As can be seen, DEA frontier depends on and is sensitive to the samplesuch that by changing the sample, the previous frontier crumbles down. Of course, not all the shortcomings of this model are due to its nonparametric nature; however, it also depends on the sample size (Bahadori et al., 2013). To overcome this problem, Simar (1996) designed a method entitled Bootstrap DEA approach to evaluate the variability of efficiency for every sampling (Löthgren & Tambour, 1996; Simar, 1996). In this method, the Bootstrap technique is used to demonstrate ranking and sensitivity of efficiency scores obtained by DEA method relative to the variation of sample combination. Bootstrap is a resampling technique proposed by Efron and Tibshirani and is applied to estimate sampling distribution characteristics of an estimator (in cases when it is difficult to obtain it with other methods) (Ebadi, 2011; Efron & Tibshirani, 1993). In the simplest format, Bootstrap is a random selecting among thousands of pseudo samples by using simple random sampling,which replace a set of series of actual samples (original sample);by using each pseudo sample a pseudo estimation of pseudo efficiency score is obtained. These thousands pseudo estimates form an empirical distribution for estimator which is used as an estimate of the sampling distribution of the main population.

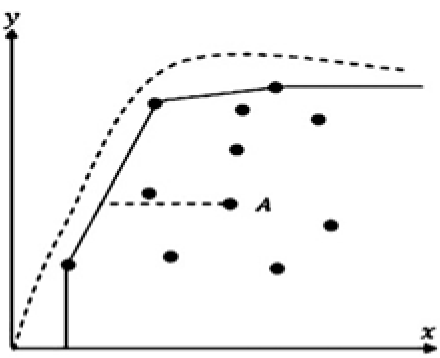

Figure 2: The first sample

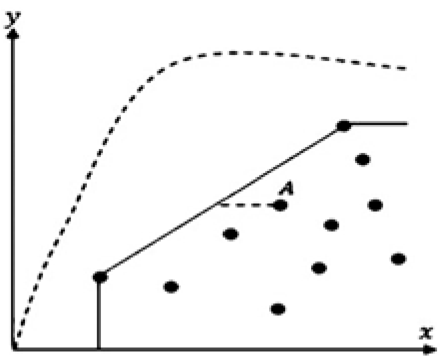

Figure 3: The second sample

Overall data generation process (DGP) and Bootstrap

The activity of a production unit produces q output $(y\in R_{+}^{q})$ using p. inputs $(x\in R_{+}^{q})$ can be shown with production set of ψ consisting of feasible and physical set (x, y), as the following equation:

$$\psi=\{(x, y)\in R_{+}^{p+q} | x\rightarrow y\} \quad \text{Input x can produces output y} \tag{1}$$

This set which may be shown as its two components, includes input set and output setso that the input set for $\forall y\in\Psi$ is shown as equation (2) and the output set for $\forall x\in\Psi$ is shown by equation (3) as below:

$$X(y)=\{X\in R_{+}^{p} | (x,y)\in\psi\} \tag{2}$$

$$Y(x)=\{Y\in R_{+}^{p} | (x,y)\in\psi\} \tag{3}$$

The relation between two sets of (2) and (3) can be explained by a set of standard assumptions provided by Shephard including the convexity assumption $X(y)$ for all y (Y(x) for all x) and the disposability assumption of inputs and outputs and so on (Shephard, 1970). The efficient frontier expressed by Farrell can be shown as a subset of $X(y)$ or $Y(x)$ which are shown as $\partial X(y)$ and $\partial Y(x)$, respectively, and in equations (4) and (5):

$$\partial X(y)=\{x | x\in X(y),\ \theta x\notin X(y),\ \forall 0<\theta<1\} \tag{4}$$

$$\partial Y(x)=\{y | y\in Y(x),\ \beta y\notin Y(x),\ \forall\beta>1\} \tag{5}$$

The already mentioned equations can be used to define as calculation indexes to obtain input- and output- oriented efficiencies for k^{th} firm (x_k, y_k) as equations (6) and (7). In the following just the input oriented model is described [2] (Färe & Grosskopf, 2006; Simar & Wilson, 1998).

$$\theta_k=Min\{\theta | \theta x_k\in X(y_k)\} \tag{6}$$

$$\beta_k=Min\{\beta | \beta y_k\in X(x_k)\} \tag{7}$$

If $\theta_k=1$, the firm (x_k, y_k) operates efficiently. Yet, if efficiency level is lower than unit $(\theta_k<1)$, this production unit (x_k, y_k) operates inefficiently and can produce using less input. For further description, it will be beneficial to definefunction level of input oriented kth firm at production level y_k, by the following equation:

$$X^{\partial}(x_k | y_k)=\theta x_k \tag{8}$$

Note that $X^{\partial}(x_k | y_k)$ is intersection point of the efficient frontier $\partial X(y)$ and radius θx_k, and in order to calculate input oriented efficiency, the ratio of radial distance of firm is to be measured (x_k, y_k) to its equivalent point on the efficient frontier of $\partial X(y)$, $X^{\partial}(x_k | y_k)$. Since the production set (ψ) and thus input set $(X(y))$ and production efficient frontier $(\partial X(y))$ are unknown, the efficiency level of k^{th} production unit (θ_k) will be unknown as well. Supposedly, using data generation process *(DGP)* as represented by ρ, we can construct a random sample set $\chi=\{(x_i,y_i) | i=1,2,\ldots,n\}$; $\hat{X}(y)$ and $\partial\hat{X}(y)$ can be obtained by using the equation method M and estimation of desired sets in the form of $\hat{\Psi}$. Therefore, efficiency of production unit (x_k, y_k) can be estimated via the following equation (Simar & Wilson, 2000).

$$\hat{\theta}_k=Min\left\{\theta | \theta x_k\in\widehat{\partial X(y_k)}\right\} \tag{9}$$

Note that sampling properties of $\hat{\Psi}$, $\hat{X}(y)$, $\partial\hat{X}(y)$ and subsequently $\hat{\theta}_k$ all depend on data generation process of ρ, which is unknown. Furthermore, even if ρ is known, achieving them by M method is very difficult, especially when M is a non-parametric method. In circumstances like thiswhere partial specification and analysis of sampling properties of estimators are very difficult or impossible, the Bootstrap method can be the most appropriate and practical for this aim. Assume data generation process of ρ is known and can be reached to an acceptable estimation like $\hat{\rho}$ by the main sample (χ), $\hat{\rho}$ is used to generate data set $\chi^{*}=\{(xi^{*}, yi^{*}) | i=1, 2, \ldots, n\}$; by using M method and this pseudo sample, an estimation of respected sets of $\hat{\Psi}^{*}$, $\hat{X}^{*}(y)$ and, $\partial\hat{X}^{*}(y)$ corresponded to the pseudo

[2] *The statements presented for input oriented model can be easily rewritten for output oriented model as well*

sample χ^*; and thus an estimation of the efficiency level $\hat{\theta}_k^*$ of k^{th} firm under the study (x_k, y_k) is obtained as relation (10).

$$\hat{\theta}_k^* = Min\left\{\theta \mid \theta x_k \in \widehat{\partial X^*(y_k)}\right\} \tag{10}$$

It should be noted that sample properties of estimators $\hat{\Psi}^*$, $\hat{X}^*(y)$, $\partial\hat{X}^*(y)$ and consequently $\hat{\theta}_k$ are totally depended on production method ρ which is an unknown method. Additionally even ρ was known, obtaining them using M method would be too difficult from the original sample χ only if $\hat{\rho}$ can be fully known; In this case, they may be difficult to calculate analytically. Yet, by using *Mont Carlo method*, an approximation of sample distributions can be simply achieved. Using $\hat{\rho}$, B pseudo samples χ_{b^*} *(b=1, 2,...B)* are generated, and then by M method, pseudo estimates of $\hat{\Psi}_b^*, \hat{X}^*(y)_b$, and $\partial\hat{X}^*(y)_b$, (b=1, 2, …, B) for each pseudo sample are obtained, and ultimately efficiency level $\{\hat{\theta}_{k,b}^*\}_{b=1}^B$ for each studied unit is calculated. The empirical density function $\{\hat{\theta}_{k,b}^*\}_{b=1}^B$ is Mont Carlo approximation of distribution $\hat{\theta}_k^*$ subject to $\hat{\rho}$. The Bootstrap method is based on the idea that if $\hat{\rho}$ is an acceptable approximation of ρ, the known Bootstrap distribution will simulate sample distribution of estimators of Ψ, $X(y)$, $\partial X(y)$ and θ_k which are of interest but unknown. Accordingly, in order to measure the efficiency level θ_k of the firm (x_k, y_k) the equation (11) must be satisfied.

$$\left(\hat{\theta}_k^* - \hat{\theta}_k\right) \mid \hat{\rho} \sim \left(\hat{\theta}_k - \theta_k\right) \mid \rho \tag{11}$$

Where, θ_k, $\hat{\theta}_k$ and $\hat{\theta}_k^*$ are defined by equations (6), (9), and (10), respectively. The above equation is valid and true if $\hat{\rho}$ is a consistent estimate of ρ. As suggested by to the equation (11), bias of $\hat{\theta}_k$ can be obtained from the main estimator of the population θ_k as follows:

$$bias_{\rho,k} = E\left(\hat{\theta}_k\right) - \theta_k \tag{12}$$

The equivalent of the above equation in the Bootstrap space may be stated as the following equation.

$$bias_{\hat{\rho},k} = E\left(\hat{\theta}_k^*\right) - \hat{\theta}_k \tag{13}$$

The expected value for $\hat{\theta}_k^*$ may be substituted by its *Mont Carlo approximation* as follows:

$$E\left(\hat{\theta}_k^*\right) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{k,b}^* = \bar{\theta}_k^* \tag{14}$$

Therefore:

$$\widehat{bias}_k = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{k,b}^* - \hat{\theta}_k = \bar{\theta}_k^* - \hat{\theta}_k \tag{15}$$

As suggested by the equation (11), one can obtain an estimate of (12) using (15). Thus:

$$bias_{\rho,k} \approx \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{k,b}^* - \hat{\theta}_k = \bar{\theta}_k^* - \hat{\theta}_k \tag{16}$$

By correcting the bias of the main estimator ($\hat{\theta}_k$), the bias-corrected estimator is obtained as equation (17) bellow:

$$\begin{aligned}\tilde{\theta}_k &= \hat{\theta}_k - bias_{\rho,k} \approx \hat{\theta}_k - \left(\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{k,b}^* - \hat{\theta}_k\right)\\ \tilde{\theta}_k &\approx 2\hat{\theta}_k - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_{k,b}^*\\ \tilde{\theta}_k &\approx 2\hat{\theta}_k - \bar{\theta}_k^*\end{aligned} \tag{17}$$

For this estimator, the term of bias-corrected estimator which is applied as the obtained bias, is not its exact value but an approximate one. Hence, the bias of estimator is not eliminated but just modified. In addition, the standard deviation of estimator $\hat{\theta}_k^*$ is shown as equation (18).

$$\widehat{se}_{\rho,k} = \left\{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_{k,b}^* - \bar{\theta}_k^*\right)\right\}^{\frac{1}{2}} \tag{18}$$

At the end of this section, the confidence

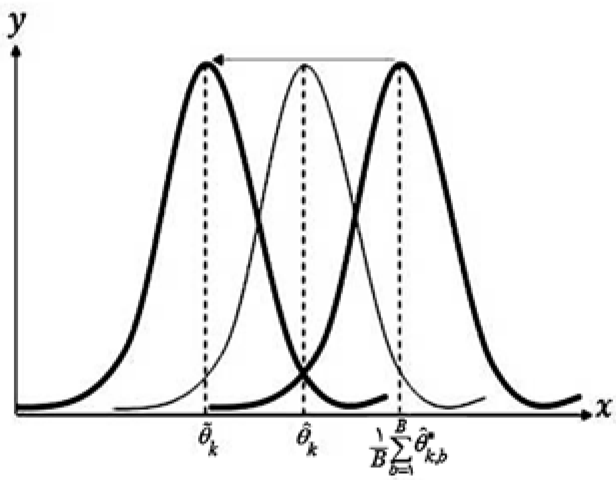

Figure 4. Bias- correction of distribution function

interval of θ_k is determined and the empirical distribution functions of $\hat{\theta}^*_{k,b}$, *(b=1, 2,..., B)* is provided after correcting the bias. We need a corrected empirical density distribution function with the centrality of bias-corrected estimator ($\tilde{\theta}_k$) from θ_k. Thus, we should move the empirical density function $\hat{\theta}^*_{k,b}$ to the left by $2\widehat{bias}_k$. As shown in figure 4, if it moves to the left by $1\widehat{bias}_k$, the empirical density function will be centralized at $\hat{\theta}_k$ rather than $\tilde{\theta}_k$ (Bahadori et al., 2013).

Therefore, the empirical density function can be shown as $\hat{\theta}^*_{k,b}$ and the confidence interval θ_k with coverage level (1-2α) can be shown by the equations (19) and (20), respectively.

$$\tilde{\theta}^*_{k,b} = \hat{\theta}^*_{k,b} - 2\widehat{bias}_k \tag{19}$$

$$\left(\hat{\theta}_{k,low}, \hat{\theta}_{k,up}\right) = \left(\tilde{\theta}_k^{*(\alpha)}, \tilde{\theta}_k^{*(1-\alpha)}\right) \tag{20}$$

In which α^{th} percent (critical value) in phrase $\tilde{\theta}_k^{*(\alpha)}$ is applied to determine confidence interval and so empirical density function $\tilde{\theta}^*_{k,b}$, *(b=1, 2,..., B)*. If the empirical density function is $\tilde{\theta}_k^{*(\alpha)}$ unbalanced, it would be preferred to select the median as the distribution centrality of $\tilde{\theta}_k$. What which has not been answered in this section is how $\hat{\rho}$ should be selected. Since the answer depends on the method of estimating *M*, the second subsection briefly describes the data envelopment analysis method (*M* method), and third one addresses various algorithms of $\hat{\rho}$ selection (Bahadori et al., 2013; Simar & Wilson, 1998, 2000).

Data envelopment analysis

The *data envelopment analysis approach* measures the efficiency of decision making unit *(x_i, y_i)* based on determination of production collection resulting from the sample $\chi = ((x_i, y_i), i=1, 2, ..., n)$ and its general form is shown as the following equation:

$$\hat{\Psi}_{DEA} = \left\{ (x_i, y_i) \in R^{p+q} \mid y_i \le Yz, x_i \ge Xz, \sum_{i=1}^{n} z_i = 1, z \ge 0, i = 1, 2, .., n \right\} \tag{21}$$

Given the above equation, the input set for the production level of *y* is estimated as follows:

$$\widehat{X(y)} = \left\{ x \in R^p_+ \mid (x, y) \in \hat{\Psi}_{DEA} \right\} \tag{22}$$

The estimated efficient input oriented frontier for the above input set at the output level of *y*, is shown by $\partial\widehat{X(y)}$; this frontier is a subset of $\widehat{X(y)}$ and obtained from definition of $\hat{\Psi}_{DEA}$. Considering the above-mentioned definitions, for each *DMU(x_i, y_i)*, the estimated efficiency level ($\hat{\theta}_i$) of i^{th}*DMU* with variable returns to scale assumption, may be computed by solving the equation (23), using the linear planning:

$$\hat{\theta}_i = \min\left\{\theta \mid y_i \le Yz, \theta x_i \ge Xz, \sum_{i=1}^{n} z_i = 1,\right. \qquad (23)$$

$$\left. z > 0, 0 \le \theta_i \le 1, i = 1,2,..,n \right\}$$

Indeed, in order to estimate the efficiency level of i^{th} firm ($\hat{\theta}_i$), the ratio of radial distance between the points in which the *DMU* operates (x_i, y_i) and its equivalent point on the efficient input-oriented frontier $(\hat{x}^{\partial}(x_i \mid y_i), y_i)$ is calculated via the above equation. In this equation, $(\hat{x}^{\partial}(x_i \mid y_i)$ is a level of consumed input that the *(DMU)* with a specific production level y_i should achieve to function efficiently. (By moving from x_i to $(\hat{x}^{\partial}(x_i \mid y_i)$ along with radius (θx_i) (Cooper et al., 2006; Schmidt, 2008).

$$\hat{x}^{\partial}(x_i \mid y_i) = \hat{\theta}_i x_i \qquad (24)$$

The Bootstrap data envelopment analysis estimator

The Bootstrap method is a statistical resampling method which it effectiveness has been proven in terms of its caliber to perform statistical inference on complex issues. The most important step in the bootstrap method is proper determination of method ρ or *DGP* from data sample of population. in The assumption underlying the use of the Bootstrap method is to estimate sample distribution of estimator using empirical distribution of estimations obtained from resampling. Bootstrap data envelopment analysis algorithms proposed by Simar, Wilson (1998) *(SW)*, and Lothgren and Tambour (1998) *(LT)*, are based on the same model of data generation process *(DGP)*, and it is assumed that to generate data at a specific level of output, the input amount required for generating pseudo sample is obtained from random radial deviations of isoquant curve of input set. Each input in the observed sample of input- output, is shown as (25):

$$(x_i, y_i) = \left(\frac{x^{\partial}(x_i \mid y_i)}{\theta_i}, y_i\right) \qquad (25)$$

Where $x^{\partial}(x_i \mid y_i) \in \partial X(y_i)$ is an unobserved point (on constructed efficient frontier) equivalent radial of firm (x_i, y_i) on the efficient frontier $\partial X(y)$ which is compared with firm location to calculate the firm efficiency. It is assumed the actual efficiency scores are taken from a similar distribution like θ_i F_θ, $i=1, 2, \ldots, n$, therefore it can be said that the Data Generation Process *(DGP)* model is an idea subject to output and input ratios and that random elements of production processes can be completely shown with the random input-oriented efficiency index. The basic idea of Bootstrap simulation is to imitate data generation process.

Bootstrap algorithms of *LT* and *SW* in each sampling are as follows: Subject to the observed ratio of inputs and outputs, resembling data are generated within two steps. In the first step, the input frontier curve is estimated using the observed sample; then using input frontier and pseudo efficiency values taken from some estimates of distribution F_θ, the amounts of Bootstrap pseudo inputs are generated by *DGP* iteration introduced in (25). This step in the algorithm provided by *LT* is established based on a simple resembling from empirical distribution of estimated efficiency scores used in generating pseudo efficiencies. Unlike the *LT* method, *SW* algorithm in the first step uses smoothed resembling process, and is based on the argument of compatibility. In the second step, the estimation of Bootstrap efficiency is performed through calculating the radial distance of efficient frontier produced by pseudo sample from pseudo input (in *LT* algorithm) or the original input amounts (in *SW* algorithm). In what follows, *LT* and *SW* algorithms are described together with their differences.

***LT* algorithm**

Steps of bootstrap algorithm provided by *LT* are as follow:

1. Using the original estimated efficiency values $\{\hat{\theta}_i, i = 1,2,\ldots,n\}$, the input-output vectors

are changed to the form of equation (26).

$$\left(\hat{x}^{\partial}(x_i \mid y_i), y_i\right) = \left(x_i.\hat{\theta}_i, y_i\right) \tag{26}$$

2. resampling is performed from *n* estimated technical efficiencies $\{\hat{\theta}_{in}\}$ and independent replacement. The efficiencies obtained from resampling are represented as δ_i^*, $i=1, 2, \ldots, n$.

3. Pseudo Bootstrap data is generated from equation (27).

$$\left(x_i^*, y_i^*\right) = \left(\frac{\hat{x}^{\partial}(x_i \mid y_i)}{\delta_i^*}, y_i\right) \tag{27}$$

4. Estimating Bootstrap efficiency scores using pseudo data is obtained by solving the model (28) with linear programming.

$$\hat{\theta}_{i,b}^{LT*} = \min_{\theta,z}\left\{\theta \mid y_i \le Yz, \theta x_i^* \ge X^* z, \sum_{i=1}^{n} z_i = 1, z > 0\right\} \tag{28}$$

5. Second to fourth steps of this algorithm are repeated in order to make B times of Bootstrap efficiency ($\hat{\theta}_{i,b}^{LT*}, b = 1, \ldots, B$) for the specific firm.

In this algorithm, the estimated Bootstrap frontier and Bootstrap efficiency scores are resembled based on the resampling technical efficiency scores obtained from empirical distribution of efficiency scores that in turn are resulted from the original sample. In addition, Bootstrap iterations of efficiency based on the resampled data are based on the original data, as the main estimations does (Lothgren, 1998; Lothgren & Tambour, 1999).

SW algorithm

The algorithm introduced by Simar and Wilson (*SW*) is different from Lothgren and Tambour's algorithm with respect to the second to fourth steps. In the second step of smoothing process, the main values of estimated efficiencies according to the empirical distribution core smoothing, are used to produce smoothed-resamples of pseudo efficiency scores. Application of smoothing process is based on reflective modification of *Gaussian kernel density* of estimation function, already mentioned by Silverman (1986). This process has been discussed by Simar and Wilson (1998) in detail. Suppose that is a non-smooth resample taken independently with replacement from empirical distribution of main values of technical efficiency $\{\hat{\theta}_{in}\}$. The smoothing process is formed in two steps: First, a small disturbance is added to δ_i^* and then the correction in resampling sequence is applied.

In order to produce smoothed pseudo efficiency first, a small disturbance equal to $h\varepsilon_i^*$ (where, *h* indicates Bandwidth and ε_i^* has been taken from a normal identical independent standard distribution) $\tilde{\delta}_i^*$ is added to δ_i^* in order to make pseudo efficiency. Considering the fact that efficiency scores are bounded within unit of distance, (the input-oriented efficiency scores resulting from *DEA*), reflective process is used in the equation (29) to generate $\tilde{\delta}_i^*$ as

$$\tilde{\delta}_i^* = \begin{cases} \delta_i^* + h\varepsilon_i^*, \text{if } \left(\delta_i^* + h\varepsilon_i^*\right) \le 1 \\ 2 - (\delta_i^* + h\varepsilon_i^*), \text{otherwise.} \end{cases} \tag{29}$$

If it is determined $\delta_i^* + h\varepsilon_i^* > 1$, $\tilde{\delta}_i^*$ changes to a symmetric image of $\delta_i^* + h\varepsilon_i^*$ reflection, on the frontier point $\tilde{\delta}_i^* = 2 - (\delta_i^* + h\varepsilon_i^*))$. One of the most important issues in use of smoothing process is selecting bandwidth parameter (*h*). As Silverman (1986) showed in his study, there are several approaches to select bandwidth (*h*). In his study, Lothgern (1998), using Mont Carlo simulation, shows that value of *h* can be calculated using a strong, automated bandwidth selection law for a variable proposed by Silverman (1986) as follows:

$$h = \circ.9 \circ n^{-\frac{1}{5}} \min\left\{\hat{\sigma}_{\hat{\theta}}, \frac{R_{13}}{1.34}\right\} \tag{30}$$

where, $\hat{\sigma}_{\hat{\theta}}$ represents anestimation of estimated internal standard deviation of efficiency values $\{\hat{\theta}_{in}\}$ and *R13* is interquartile range of empirical

Table 1
Summary of Statistics

	Date(kg)	Date palm(N)	Water(h)	Labor(p/d)	Land(h)
Maximum	10000	500	3085	103	197100
Mean	2241.96	170.34	540.46	28.22	21152.14
Minimum	300	10	25	2	1230
Variance	3842300	19885.94	380502.5	472.1751	903295263
Standard deviation	1960.179	143.1673	627.9034	22.12528	30611.141

distribution of $\{\hat{\theta}_{in}\}$. After above steps, smoothed resampling efficiency values (γ^*) is determined via modifying $\tilde{\delta}_i^*$ as follows:

$$\gamma_i^* = \frac{\bar{\delta}_i^* + \left(\tilde{\delta}_i^* - \bar{\delta}_i^*\right)}{\dfrac{\sqrt{1+h^2}}{\hat{\sigma}_{\hat{\theta}}^2}} \tag{31}$$

where, $\bar{\delta}_i^* = \sum_{i=1}^{n} \frac{\delta_i^*}{n}$ is resampling main efficiency scores mean value. The second fundamental difference between the two algorithms provided by *LW* and *SW* is in the fourth step. In *SW* algorithm, the estimated Bootstrap input-oriented efficiency scores of i^{th} production firm is obtained according to the ratio of radial distance (at fixed output level) of consumed input of i^{th} firm; resulting from original data to its corresponding point located on Bootstrap pseudo production is quant curve, which is in contrast with *LT* algorithm in which the pseudo efficiency scores are produced from the pseudo data associated with Bootstrap pseudo frontier.

SW algorithm can be briefly described as below:

1. Input-output vectors used in the calculation of main efficiency scores (calculated from the original sample) $\{\hat{\theta}_i, i=1,2,\ldots,n\}$, are replaced into the following equation:

$$\left(\hat{x}^{\partial}\left(x_i, y_i\right), y_i\right) = \left(x_i \hat{\theta}_i, y_i\right) \tag{32}$$

2. In this stage, the smoothed resampling efficiency scores (γ^*) are as follows:

2.1 Using the equation (30), estimated efficiency scores are applied to determine bandwidth (h).

2.2 The values of $\{\delta_i^*\}$ are produced by resampling along with replacement from empirical distribution of estimated efficiency scores $\{\hat{\theta}_{in}\}$.

2.3 Using equation (29), a string of $\{\tilde{\delta}_i^*\}$ is produced.

2.4 Bootstrap pseudo data are generated as the following equation:

$$\left(x_i^*, y_i^*\right) = \left(\frac{\hat{x}^{\partial}\left(x_i, y_i\right)}{\gamma_i^*}, y_i\right) \tag{33}$$

4. The estimator of Bootstrap efficiency scores is obtained using pseudo data and by solving the model (34) with linear planning:

$$\hat{\theta}_{i,b}^{SW*} = \min_{\theta,z}\left\{\theta \mid y_i \le Yz, \theta x_i \ge X^* z, \sum_{i=1}^{n} z_i = 1, z > 0\right\} \tag{34}$$

5. The second to fourth steps of this algorithm are repeated B times to make a B-member set of Bootstrap efficiencies $(\hat{\theta}_{i,b}^{LT*}, b=1,\ldots,B)$ of the specified firm (Lothgren, 1998; Silverman, 1986;Simar & Wilson, 1998).

Data and Specification of Variables

With about 100 palm groves and more than one-hectare cultivation, RoodkhanehbarRegion (including the villages of MiyanChilan, Bagh-Narges, Regab, Hizbandegan, Delbanan, Chah-Saifullah, and Baghshah) is one of the date producers of Hormozgan Province. Among the varieties planted in Hormozgan Province, the Keriteh variety is the dominant palm,and accordingly, in this study, the efficiency of Keriteh palm production has been evaluated. Random

Table 2

Bias- Corrected Technical Efficiency Scores and the Formation of Confidence Intervals

P- G	Es-Bi	Bi-Corr	S-Dev	Conf-int		Es-Bi	Bi-Corr	S-Dev	Conf-int		S-E
	CRS			L	U		VRS		L	U	
1	0	1	0	1	1	0	1	0	1	1	1
2	-0.0006	0.0703	0.0055	0.0607	0.0786	-0.0291	0.1274	0.0058	0.0601	0.0793	0.551805
3	-0.0002	0.073	0.0061	0.0628	0.0827	-0.0045	0.0808	0.006	0.0622	0.082	0.903465
4	0.0005	0.0662	0.0075	0.0551	0.0786	-0.0282	0.1929	0.0115	0.1199	0.1575	0.343183
5	0	0.0268	0.0028	0.0224	0.0311	-0.0618	0.1504	0.0028	0.0223	0.0311	0.178191
6	0.0001	0.0133	0.0016	0.0109	0.0157	-0.0534	0.12	0.0015	0.011	0.0158	0.110833
7	-0.0003	0.0669	0.0058	0.0567	0.0756	-0.017	0.1141	0.0062	0.07	0.0903	0.586328
8	-0.0002	0.035	0.0032	0.0296	0.0396	-0.0825	0.1995	0.0031	0.0297	0.0397	0.175439
9	0.0002	0.1336	0.015	0.1108	0.1564	-0.2659	0.6656	0.015	0.1104	0.1564	0.200721
10	-0.001	0.1868	0.0162	0.1594	0.2115	-0.0957	0.3794	0.0154	0.1611	0.212	0.492356
11	0	0.1537	0.0164	0.1278	0.1786	-0.034	0.2215	0.0166	0.1273	0.1791	0.693905
12	0	1	0	1	1	0	1	0	1	1	1
13	-0.0016	0.2122	0.0172	0.1806	0.2379	-0.7914	1	0.0176	0.1802	0.2396	0.2122
14	-0.0005	0.1138	0.0137	0.0953	0.1376	0.0002	0.1161	0.0137	0.0949	0.1376	0.980189
15	-0.0002	0.0853	0.0099	0.0696	0.1004	-0.209	0.5031	0.01	0.0697	0.1005	0.169549
16	-0.0001	0.018	0.0016	0.0153	0.0205	-0.0775	0.1727	0.0016	0.0153	0.0204	0.104227
17	-0.0014	0.2967	0.0249	0.2566	0.3383	-0.1297	0.5543	0.0244	0.2548	0.3372	0.53527
18	-0.0011	0.2971	0.0308	0.2477	0.3422	-0.705	1	0.031	0.2467	0.3423	0.2971
19	0.0002	0.0932	0.0109	0.0763	0.1104	-0.106	0.306	0.0108	0.0769	0.1106	0.304575
20	-0.0002	0.181	0.0214	0.1491	0.2151	0.0004	0.1814	0.0212	0.1487	0.2145	0.997795
21	-0.0012	0.2586	0.0213	0.2207	0.2923	-0.1567	0.5936	0.0196	0.2474	0.3118	0.435647
22	-0.0023	0.2642	0.0231	0.2226	0.2978	-0.1589	0.6097	0.0197	0.2581	0.3247	0.433328
23	-0.0017	0.2553	0.0222	0.2169	0.2888	-0.1462	0.5541	0.0208	0.2246	0.2919	0.460747
24	-0.0008	0.1206	0.01	0.1028	0.136	-0.0461	0.2158	0.0108	0.1059	0.1408	0.558851
25	0.0024	0.3147	0.0347	0.262	0.371	-0.2429	1	0.0114	0.7396	0.7753	0.3147
26	-0.0006	0.4087	0.0331	0.3543	0.4646	-0.1095	0.6544	0.0345	0.378	0.4928	0.624542
27	0	1	0	1	1	0	1	0	1	1	1
28	-0.0006	0.2165	0.0207	0.1827	0.249	-0.0323	0.2849	0.0182	0.1909	0.2504	0.759916
29	-0.0011	0.2712	0.0253	0.2298	0.311	-0.0566	0.4494	0.0255	0.2961	0.3799	0.603471
30	-0.0006	0.2707	0.0253	0.2297	0.3084	-0.0402	0.3563	0.0225	0.2372	0.3117	0.759753
31	-0.0012	0.1834	0.0173	0.1544	0.2086	-0.101	0.3825	0.017	0.1539	0.2077	0.479477
32	-0.0001	0.1297	0.012	0.1102	0.148	-0.1961	0.5218	0.0121	0.1101	0.1487	0.248563
33	-0.0008	0.1359	0.0125	0.1146	0.1542	-0.2714	0.6769	0.0125	0.1141	0.1535	0.200768
34	-0.0006	0.1348	0.0134	0.1125	0.1546	-0.1162	0.3664	0.0136	0.1126	0.1546	0.367904
35	-0.0019	0.2511	0.0234	0.2109	0.2842	-0.0539	0.4164	0.0239	0.2724	0.3519	0.603026
36	0.0002	0.3237	0.0307	0.2755	0.3725	-0.0694	0.5406	0.0303	0.3549	0.4537	0.598779
37	-0.0046	0.5443	0.0515	0.457	0.6163	-0.1049	1	0.0221	0.8546	0.9241	0.5443
38	-0.003	0.4347	0.042	0.3635	0.4972	-0.0629	0.5682	0.0371	0.3816	0.5017	0.765048
39	-0.0007	0.1673	0.0191	0.1366	0.1965	-0.0835	0.3335	0.019	0.1369	0.196	0.501649
40	-0.0009	0.1324	0.0142	0.1083	0.1528	-0.0431	0.3012	0.0176	0.1886	0.2455	0.439575
41	-0.0004	0.3403	0.0346	0.2857	0.392	-0.0672	0.4956	0.0266	0.3173	0.4047	0.686642
42	0	0.018	0.0021	0.0147	0.0212	-0.0682	0.1544	0.0021	0.0147	0.0213	0.11658
43	-0.0002	0.0449	0.0047	0.0372	0.0519	-0.1079	0.2603	0.0046	0.0371	0.0513	0.172493
44	0.0001	0.0943	0.0109	0.0776	0.1116	-0.0883	0.2711	0.0106	0.0775	0.111	0.347842
45	-0.0004	0.1188	0.0121	0.0992	0.1371	-0.0093	0.1367	0.0121	0.0993	0.1369	0.869056
46	-0.002	0.3041	0.0275	0.2567	0.3434	-0.0674	0.4999	0.0279	0.319	0.4122	0.608322
47	-0.0023	0.2061	0.017	0.1744	0.2311	-0.1138	0.4398	0.0172	0.1825	0.2385	0.468622
48	-0.0005	0.1152	0.0105	0.0974	0.1311	-0.0287	0.2098	0.0125	0.1337	0.1741	0.549094
49	-0.0002	0.0821	0.0077	0.0695	0.0937	-0.1584	0.3985	0.0078	0.0697	0.0943	0.206023
50	-0.0001	0.1769	0.0156	0.1517	0.2022	-0.0411	0.286	0.016	0.1776	0.2305	0.618531
Average	- 0.00065	0.2288 28	0.0161	0.2023 16	0.2533 98	-0.11064	0.4412 6	0.0147 96	0.2329 42	0.2807 48	0.503608 8

sampling and questionnaire were used to determine the sample size of the studied population, and Bartlett's table was used to collect data. The Table was applied to determine minimum returned sample size of a given population size for continuous and categorical data. Based on this table, for a population of 100 palm groves over a hectare, at probability level of 10% error,

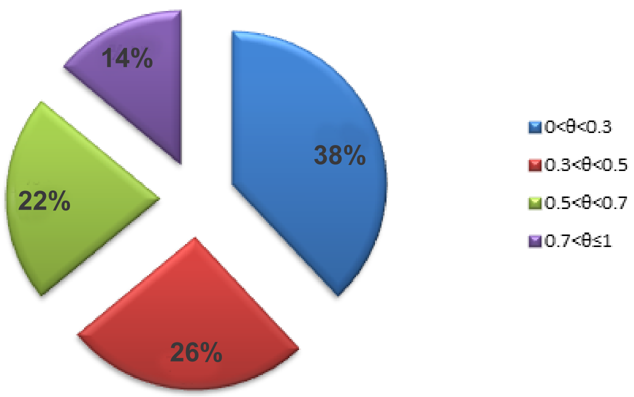

Figure 5. Distribution of pure technical efficiency scores for date producers

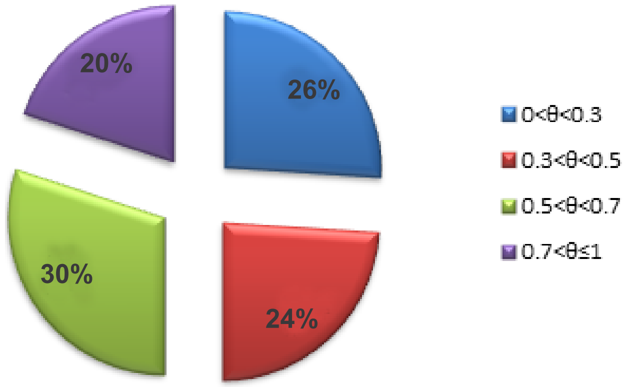

Figure 6. Distribution of pure scale efficiency scores for date producers

a sample with a population of at least 46 palm groves should be collected (Barlett et al., 2001). As such, in this paper, from producer population with over a hectare cultivation of palm in Roodkhanehbar region in 2013, a sample with the size of 50 palm groves was taken. Information regarding the amount of palm product (Kilograms) and four basic inputs including labor (person/Labor Day), land (hectares), water (h), and the number of date palms (N) were used to evaluate the efficiency of palm production in the region. A summary of the statistics is listed in Table 1.

RESULTS AND DISCUSSION

Using smoothed Bootstrap, the input-oriented data envelopment analysis provided by Simar and Wilson, the technical efficiency scores and related confidence intervals with 1000 iterations, (B=1000) at 95% confidence level were calculated in this survey, assuming VRS[3] and CRS[4]. CRS assumption is only appropriate when all firms are operating at an optimal scale, and it implies the total technical efficiency (TE). Using CRS specification when all firms are not operating at an optimal scale results in gross measures of TE by scale efficiencies (SE). Additionally, using VRS specification permits the calculation of TE devoid of these SE effects and refers to pure technical efficiency. Therefore, TE is decomposed to pure efficiency and scale efficiency; therefore, scale efficiency can be calculated by the ratio of technical efficiency and pure technical efficiency scores.

The results are listed in Table 2 and Figure 2. The bandwidth used in this study was computed via the equation (30) proposed by Silverman and was equal to 0.0005.

Abbreviations used in the above table are as listed:

P-G: Palm groves

Es-Bi: Estimated bias

Bi-Corr: Bias-corrected technical efficiency

S-Dev: Standard Deviation

Conf-Int (L-U): Confidence interval (lower-bound-upper bound)

S-E: Scale Efficiency

As can be seen in the table above, the average pure technical efficiency scores resulting from Bootstrap input-oriented DEA under VRS for Keriteh palm cultivation were about 44%, which indicates that growers in this region are able to produce the same amount of product with about an average of 66% saving in resource consumption, and therefore, reduce their production costs. Figure 5 also demonstrates that 38% of studied palm groves had efficiency less than 30% and that only 14% of growers operated efficiently.

---

[3] Variable Returns to Scale

[4] Constant Return to Scale (here, restrictions [$\sum_{i=1}^{n} z_i = 1$] in CRS model is not applied.)

Similarly, as shown in the Figure 6 below, the scale efficiency score for only 20% of date producers was more than 0.7. This shows that the activity of most of them has been far distant from efficient production capacity.

CONCLUSION AND RECOMMENDATIONS

Since producing dates is either directly or indirectly the main activity and source of income for people in Roodkhanebar Region, any increase in technical efficiency (TE) of the activity leads to saving product inputs consumption, reduces production costs, and increases output level with existing inputs, and finally, increases both profit and welfare of the people of the region.

Due to the importance of the issue, this study aimed to evaluate the management (pure) and scale efficiency of date production in this region. As shown in Table 2, date producers of this region, having average management (pure) and scale efficiencies scores equal to 0.44 and 0.50 respectively, operated weakly, and significant savings will be possible for given quantity of outputs by efficient management and operating in optimum scale. Furthermore, as listed in Figures 2 and 3, the distribution of scores of management and scale efficiencies showed 64% of producers had management efficiency less than 0.50 and 50% of producers had scale efficiency less than 0.50. Given the low efficiency scores and high number of inefficient producers (in terms of both management and optimum scale efficiencies), and alsobecauseincreasing lower levels of efficiency is more convenient and easier than increasing it for higher levels, (e.g., increasing efficiency from 20 to 50% is easier than increasing it from 70 to 100%), the efficiency of producers in this region can be remarkably increased by taking appropriate promoting and educating actions to improve the general knowledge of optimum resource consumption andto facilitate proper input access. In this regard, it is recommended to establish a center for training and experience share, where successful orchardists could share their experiences with others in an attempt to optimize efficient use of inputs and production as well.

ACKNOWLEDGEMENT

We would like to take this opportunity to thank Mr. Omid Tajik who helped us withdata collection and Ms. Pooneh Khajeh Hassani for editing the final version of the paper.

REFERENCES

Bahadori, A., Hosseini Nahad, S., & Habibinia, G. (2013). Use of Bootstrap simulation process for estimating non parametric efficient frontier “Investigating problems existing in the process provided in Saeed Ebadi’s paper. *Research in Operation And Its Applications, 37*(2), 113-135.

Balcombe, K., Fraser, I., Latruffe, L., Rahman, M., & Smith, L. (2008). An application of the DEA double bootstrap to examine sources of efficiency in Bangladesh rice farming. *Applied Economics, 40*(15), 1919-1925.

Barlett, James E, Kotrlik, Joe W., & Higgins, Chadwick C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal, 19*(1), 43-50.

Brümmer, B. (2001). Estimating confidence intervals for technical efficiency: the case of private farms in Slovenia. *European Review of Agricultural Economics, 28*(3), 285-306.

Cooper, W., Seiford, L., & Tone, K. (2006). *Introduction to data envelopment analysis and its uses: with DEA-solver Software and References*: Springer Science and Business Media.

Dong, Fengxia, & Featherstone, Allen M. (2006). Technical and scale efficiencies for chinese rural credit cooperatives: a bootstrapping approach in data envelopment analysis. *Journal of Chinese Economic and Business Studies, 4*(1), 57-75.

Ebadi, S. (2011). A method for ranking efficiency scores using Bootstrap. *Research in Operation and its Applications, 29*(2), 29-44.

Efron, B., & Tibshirani, R. (1993). An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, Vol. 57. *New York and London: Chapman and Hall/CRC*, 321-335.

FAO, STAT. (2012). FAOSTAT-Statistical Database, 2012.

Färe, R., & Grosskopf, . (2006). *New directions:*

efficiency and productivity (Vol. 3): Springer Science and Business Media.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253-290.

Gocht, Al., & Balcombe, K. (2006). Ranking efficiency units in DEA using bootstrapping an applied analysis for Slovenian farm data. *Agricultural Economics, 35*(2), 223-229.

Lothgren, M. (1998). How to bootstrap DEA estimators: a Monte Carlo comparison. *WP in Economics and Finance* (223).

Lothgren, M., & Tambour, M. (1999). Testing scale efficiency in DEA models: a bootstrapping approach. *Applied Economics, 31*(10), 1231-1237.

Löthgren, M., & Tambour, M. (1996). *Scale Efficiency and Scale Elasticity in DEA-models: A Bootstrapping Approach*: Economic Research Inst.

Odeck, J. (2009). Statistical precision of DEA and Malmquist indices: a bootstrap application to Norwegian grain producers. *Omega, 37*(5), 1007-1017.

Schmidt, S. (2008). *The measurement of productive efficiency and productivity growth:* Oxford: Oxford University Press.

Shephard, RW. (1970). *The theory of cost and production functions, Princeton.* New Jersey: Princeton Univ. Press (1st edition, 1953).

Silverman, B. (1986). *Density estimation for statistics and data analysis* (Vol. 26): CRC press.

Simar, L. (1996). Aspects of statistical analysis in DEA-type frontier models. *Journal of Productivity Analysis, 7*(2-3), 177-185.

Simar, L., & Wilson, P. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44(1), 49-61.

Simar, L., & Wilson, P. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics, 27*(6), 779-802.

How to cite this article:

Zamanian, G., & Khajeh Hassani, M. (2016). Evaluation of Palm groves technical efficiency using bootstrap data envelopment analysis: A case study of Roodkhanehbar area, Iran. *International Journal of Agricultural Management and Development, 6(4), 475-487.*

URL: http://ijamad.iaurasht.ac.ir/article_526565_9bca2075558aca055e9ebc7c2be8b9e5.pdf