Resource Efficiency Estimation and Digital Recommendation: Sustainable Pathway to Improve Paddy Farmers' Productivity

Piyush Kumar Singh*[1], Shiladitya Dey[1], Anirban Pal[1]

[1] Centre for Rural Development and Innovative Sustainable Technology, Indian Institute of Technology Kharagpur, India

Contributed Paper prepared for presentation at the 97th Annual Conference of the Agricultural Economics Society, University of Warwick, United Kingdom

27 – 29 March 2023

*Piyush Kumar Singh
Indian Institute of Technology Kharagpur, West Bengal, India, 721302.
psingh@iitkgp.ac.in

Acknowledgments: This research was supported by Indian Council of Social Science Research, IMPRESS Scheme, Ministry of Education, Government of India

Abstract

The study assesses the resource use efficiency of smallholder paddy farmers with/without considering undesirable outputs through the mobile-based application. Further, the study performs an impact assessment of digital recommendations on farmers' paddy yield improvement. A mobile app-based questionnaire was used to collect data from 153 paddy farmers in eastern India. The study employed Data Envelopment Analysis (DEA) to identify the farmers' resource use efficiency with/without undesirable output. We found lower farm eco-efficiency scores with undesirable output in the model compared to the case of not considering the undesirable output analysis. Results also showed that farmers are over-utilizing fertilizers, farming machinery, and labor in farming, which needs to be reduced to the recommended optimal level. Finally, using the Propensity Score Matching (PSM), we observed that the farmers achieved better paddy yield, i.e., an additional 0.6t/ha paddy, due to the adaptation of mobile-based recommendations. Subsequently, we used probit modeling to estimate the critical factors for adopting mobile-based services. Results show that farmers' education level, farm experience, social capital, and market information play a significant role in mobile-app-based recommendation adoption. This study supports that farmers need to be suggested to use digital advisory services, and state/central policies may be aligned towards strengthening farmers' capacities for applying digital services in the farming system.

Keywords Eco-efficiency estimation, Mobile application based input recommendation, paddy yield improvement, Slack-Based Data Envelopment Analysis, Propensity Score Matching Approach

JEL code Q1, Q12, Q15, Q16

1. Introduction

Rice is the most widely produced cereal crop in the world. Rice is the staple food for 3.5 billion people worldwide (IRRI, 2013). For nearly 520 million poor people in Asia, rice provides 50% of their daily calorie needs and a substantial part of their protein requirement (FAO, 2012). Rice is also the primary source of employment and income for more than 200 million households across developing countries in the world (FAO, 2004). Currently, 85% (408 MMT) of milled rice is utilized for human consumption. India and China account for 50% of the world's milled rice consumption.

In the past century, the global population has increased fourfold and is likely to upsurge by over one billion in the next 15 years, reaching almost 8.5 billion by 2030 and 11.2 billion by 2100 (United Nations, 2015). With the world's population growth, rice production has to be improved proportionately to avoid a food crisis (Carlsson-Kanyama & González, 2009). Nearly 870 million people in developing countries where rice is responsible for food security are estimated to suffer from chronic undernourishment (Muthayya et al., 2014). Therefore, to maintain food security, rice yield needs to be doubled by 2050 (Ray et al., 2013). Farmers often overuse inputs like seeds, irrigation water, chemical fertilizer, and biocides (insecticides, fungicides, and herbicides) to maximize the rice yield (Mobtaker et al., 2010). The increased amount of input used to improve the grain yield also enhances greenhouse gas emissions and minimizes profit (Erdal et al., 2007). Therefore, using agricultural inputs at an optimum rate is crucial to maximizing rice production's profit and long-term sustainability (Banaeian et al., 2012; Chandio et al., 2020; Yuan & Peng, 2017).

In developing nations, several researchers studied paddy-producing farmers' resource use efficiency to make their production system efficient in terms of yield and profit maximization (Tiongco & Dawe, 2002). Most of these research works used Data Envelopment Analysis (DEA) as a suitable efficiency measurement tool (Amid et al., 2016; Banaeian & Zangeneh, 2011; Chauhan et al., 2006; Hosseinzadeh-Bandbafha et al., 2017; Houshyar et al., 2012). This approach can include multiple inputs and outputs to evaluate the efficiency and derive the efficient frontier. DEA is a relative measurement that generates efficiency scores for all the Decision-Making Units (DMUs) in a given process against the best-operating units. The DMUs on the frontier are the most efficient, bearing a score of 1, and the remaining units below the frontier are inefficient units bearing scores of less than 1. In the past, using this nonparametric tool, several studies identified the resource use efficiency of farmers involved in different crop production and recognized those inputs which played a significant role in efficiency gain. However, based on the findings, any study rarely provides input recommendations to inefficient farmers and monitors whether farmers' yields improved following the experts' recommendations. Hence, we have selected two objectives to fulfill the research gaps. The first objective was to estimate paddy producers' resource-use efficiency and recommend the optimum quantity of agricultural inputs to the inefficient farmers. Another aim was to measure the impact of experts' recommendation adaptation on paddy yield.

2. Methodology

2.1 Sampling and data collection

This study was carried out in the state of West Bengal of India. In this eastern Indian state, smallholding agriculture is the primary source of income for a larger part of the population. West Bengal accounts for the largest paddy production in India. Paddy production needs many input resources as it is a resource-intensive crop. Thus, optimizing the input consumption to avert the over-use of resources in paddy production is essential. The study follows a multi-stage random sampling, including data collection from Darjeeling district villages. During the COVID-19 pandemic, we developed a mobile application to collect farmers' data and avoid the face-to-face questionnaire survey. We installed the app on 200 farmers' smartphones and asked them to fill in the questions in the form. Out of the total, 155 farmers' successfully filled the app-based questionnaire. Data obtained from these farmers were used for efficiency estimation. Then, we calculated the efficiency and, based on that, we provided input recommendations to the inefficient farmers to improve their yield.

2.2 Data Envelopment Analysis (DEA)

The efficient use of a farm's limited energy resources is a significant concern for policymakers worldwide. As a result, this study focuses on calculating the resource-use efficiency of paddy production in eastern India and recommends the optimum quantity of agricultural inputs. In this research, we propose a DEA-based model to estimate efficiency as a function of land size, land rent, human labor, machines, and fertilizers. Irrigation energy costs have been left out purposefully because agriculture in the study area is primarily rainfed (Baruah et al., 2004). Additionally, undesirable outputs in the form of GHG emissions have also been considered to obtain a more accurate measure of efficiency. Various mathematical programming techniques can be used to assess production efficiency. Data Envelopment Analysis (DEA) is a nonparametric mathematical programming approach that can determine the relative efficiency of decision-making units (DMUs) (Sengupta, 1987). DMUs, in this case, would refer to prospective farmers engaging with the digital platform. DEA calculates the technical efficiency (TE) of a group of DMUs by comparing them to the most efficient ones, within the group, in terms of input utilization and output production. Compared to other parametric frontier-based approaches, DEA is not dependent on a pre-determined production function, which is a significant advantage when comparing agricultural productivity (Malana & Malano, 2006). The frontier function in DEA comprises weighted combinations of observed most efficient DMUs (that form the frontier); the TE of inefficient DMUs can then be estimated as their relative distance from the efficient frontier (Nghiem et al., 2007).

TE measures the capacity of a DMU to generate maximum outputs using a prescribed amount of inputs. As is known, agricultural activity hardly ever runs under perfect market conditions, where higher input consistently correlates to improved yield. Hence, the DEA methodology described here uses variable returns to scale (VRS) to compute efficiency. The input orientation method is applicable in the case of agriculture as it determines which of the inputs can be lowered without affecting the output. Therefore the current study has taken an input-oriented approach in line with similar studies (Mousavi-Avval et al., 2011; Phadnis & Kulshrestha, 2012; Soni et al., 2018) to discover variables that encourage greater yields. In agricultural production systems, outputs can be either desirable (good) or undesirable (bad). Figure 1 is an example of a frontier function for two outputs and one input system, one good and the other bad. The axes show the proportions of various outputs to the given input. The efficient frontier's convex shape is related to the input-oriented approach's requirements. As is observed, the efficiency frontier comprises O, B, C, D, E, and F, efficient farmer DMUs, while p1 and p2 are inefficient farmers. DEA can also determine the weight of the efficient DMUs that contribute to the virtual DMU that serves as the reference for each inefficient DMU and the slacks in each inefficient DMU's particular inputs and outputs. The slack analysis can help us understand how inefficient farmers can increase their efficiency by making proportionate input management decisions.

(Insert Figure 1 here)

The effect of undesirable outputs, such as GHG emissions on farm efficiency, cannot always be measured directly. As a result, shadow values of undesirable outputs are calculated using non-radial measures in a nonparametric model. Chambers et al. (1998), Kwon & Lee (2015), and Barra & Zotti (2016) are proponents of this paradigm. Nonparametric techniques like DEA have several advantages over parametric methods. For instance, data outliers are more noticeable when using a nonparametric technique (Zhou et al., 2018). Furthermore, even with a few observations, we can calculate several parameters accurately (Emrouznejad et al., 2009). The lack of a pre-determined functional form makes nonparametric techniques more adaptable and simple to apply. Even the shadow value of unwanted outputs can be estimated using nonparametric techniques. In contrast, the deterministic approach ignores random errors, and the stochastic approach fails to satisfy the monotonicity criterion (Zhou et al., 2018).

2.3 Undesirable Outputs

Undesirable outputs, unlike standard outputs, hurt the environment. Thus, a radial DF cannot adequately reflect the goal of reducing favorable outputs while simultaneously raising undesirable outputs, as both good and bad outputs change in the same proportion. These issues have been solved thanks to non-radial measures. The Directional Distance Function (DDF) technique is a non-radial measure that was first introduced by Tone (2001). A DDF works directly with the slacks in production factors, reducing both inputs and unwanted outputs while increasing desirable outputs. Regarding separating efficiency scores, DDF outperforms radial measures (Barra & Zotti, 2016). The DDF can measure differently based on the chosen directional vector. Therefore, DDF based on endogenous directional vectors provides a better valuation of the impact of undesirable outputs on TE.

2.4 Slacks-based DEA

DDF has been vastly used in prior literature (Barra & Zotti, 2016; Chambers et al., 1998; Flückiger & Vassiliev, 2007) to determine the shadow price of environmental inefficiencies and unwanted outputs. Therefore, DDF has been paired with DEA in the present study to estimate the impact of undesirable outcomes on farm efficiency in the form of GHG emissions. The system is based on production technology, wherein we consider a selection of $j = 1, \ldots, J$ farmers who use an input vector, x, to generate desired output vector, y, and an undesirable output vector, b. A general representation of the Production Possibility Set (PPS), which meets the basic axioms of a convex, closed, non-empty set, has been shown in Eq. (1):

$$Y(x) = \{ (y, b) | \, x \; produce \; (y, b), x \in \Re_+^M, y \in \Re_+^{S_1}, b \in \Re_+^{S_2} \} \tag{1}$$

where $Y(x)$ represents the Production Possibility Set (PPS); $y = (y_1, \ldots, y_{S_1})$ are the set of desirable or good outputs while $b = (b_1, \ldots, b_{S_2})$ are the set of undesirable or bad outputs that are produced by a set of inputs, $x = (x_1, \ldots, x_m)$. Then, a DDF for the specified PPS can be defined as follows:

$$\overrightarrow{D_0}\left(x, y, b; g_y, -g_b\right) = \sup\{\beta: (y + \beta g_y, b - \beta g_b) \in Y(x) \tag{2}$$

where the directional vector $(g_y, -g_b)$ is projected by any efficient DMU forming the frontier. In order to calculate this distance function, a nonparametric DEA approach has been adopted, and the environmental constraints have been defined as follows:

$$\overrightarrow{D_0}\left(x, y, b; g_y, -g_b\right) = \sup\{\beta: (y + \beta g_y, b - \beta g_b) \in Y(x) \tag{3}$$

$$s.t. \sum_{j=1}^{J} \lambda_j \times y_j \geq (1 + \beta) \times y_j \tag{4}$$

$$\sum_{j=1}^{J} \lambda_j \times b_j = (1 - \beta) \times b_j \tag{5}$$

$$\sum_{j=1}^{J} \lambda_j \times x_{mj} \leq x_{mj};\ m = 1, \ldots, 6 \tag{6}$$

$$\beta \geq 0, \lambda_j \geq 0\ \forall\ j = 1, \ldots, J \tag{7}$$

where the j index represents the farmer being evaluated, and their efficiency score is indicated by β. $g_j = (-b_j, y_j)$ is considered the directional vector, such that the minimum drop in undesirable output and the maximum rise in desirable output can happen simultaneously. Furthermore, the intensity vector (raw weights representing the proportion of other efficient farmers needed to maximize efficiency) is expressed by the sign of λ. Evidently, the ratio of shadow values for desirable and undesirable outputs can be estimated from the values of either the efficient farmers or the projection of ineffective farmers onto the PPS (as shown in Fig. 1).

2.5 Impact estimation

2.5.1 Problems associated with impact evaluation

There may be several theoretical explanations why (application) might improve paddy yield, but how can we confirm that better productivity of app adopters compared to non-app adopters is due to mobile app adoption (or not)?

Experimental data is suitable for a causal inference between mobile app adoption and yield improvement, as it can resolve the *missing data* problem. However, observational household data used for this study suffers from a *missing data* problem. Another issue with household data is *self-selection*, i.e., households determine whether they adopt advanced technology, and their decision may be correlated to the welfare originating from advanced technology adoption. It signifies that technology adoption and welfare are probably a two-way relationship whereby adopting advanced technology can help achieve welfare, such as an increase in income, educational status, and health – which may foster the advanced technology adoption.

2.5.2 Estimation strategy & empirical model

If a mobile app was randomly installed on farmers' mobile – as it would be in the experimental approach – we could estimate the causal effect of that mobile app on farm productivity as the difference in average yield between users and non-users of the mobile app. Yet, with observational data, the study needs to incorporate some statistical solutions to such a critical matter of causal inference.

The study may refer to an approach to defining paddy yield and mobile-app adoption as follows:

$$Y_i^M = D^M(X_i) + \varepsilon_i^M \qquad M = 0, 1, \tag{8}$$

$$M_i = A(H_i) + \delta_i, \tag{9}$$

Where Y_i^M denote paddy yield of household i that adopts the new mobile-app M. Thus, Y_i^1 and Y_i^0 would denote the paddy yield of household i in case the farm-household adopts or does not adopt the mobile app, respectively. Paddy productivity depends on a vector of some observed variables X_i and on a vector of unobserved variables, ε_i^M. M_i is a binary variable equal to 1 if the farmer adopts the mobile app and 0 otherwise. H_i is a subset of X_i and incorporates observed factors influencing the choice to avail of the mobile app, while the random variable summarizes other unobserved variables related to farm-household δ_i.

Household attributes concerning mobile-app adoption can be the outcome of the decision-making process whereby the standard separability condition between production and consumption does not hold. The production decision is influenced by some household characteristics that influence paddy yield.

Does mobile-app adoption improve paddy yield, or is the positive correlation between the two because households with better yield are richer enough to use a smartphone, internet connectivity, and adopt mobile-app? In other words, we are interested in identifying the correlation between mobile-app adoption and paddy yield and the underlying causation (Becker & Ichino, 2002).

According to Rosenbaum & Rubin (1983), in a counterfactual framework, we can estimate the causal effect as the *average treatment effect* as

$$\omega = E(Y_i^1 - Y_i^0). \tag{10}$$

In the causal effect estimation, a fundamental problem arises as we observe either Y_i^1 or Y_i^0 and not both for each household. Hence, our observation can be written as

$$Y_i = M_i Y_i^1 + (1 - M_i) Y_i^0 \quad M = 0, 1. \tag{11}$$

We can rewrite Eq. (4) for ω as

$$\omega = P\,.\,[E(Y^1 \mid M = 1) - [E(Y^0 \mid M = 1)] + (1 - P)\,.\,[E(Y^1 \mid M = 0) - [E(Y^0 \mid M = 0)] \tag{12}$$

Where P is the probability of observing a household with $M = 1$ in the dataset, the above equation indicates that the impact of mobile-app adoption for the whole dataset is the weighted average of the impact of mobile-app adoption between the two groups of farm-households. Those currently using the mobile app or *treated* and those not using the mobile app or *controls* are weighted by their relative frequency. Still, we are unable the measure the unobserved counterfactuals $E(Y^1 \mid M = 0)$ and $E(Y^0 \mid M = 1)$, which is the major issue with *causal inference* (Heckman et al., 1998).

If the mobile app was randomly installed on the farmers' mobile, we could substitute the unobserved counterfactuals, $E(Y^1 \mid M = 0)$, with the actual paddy yield $E(Y^1 \mid M = 1)$ as the two would be equal or close to equal. However, mobile-app adoption is not random. Hence, there is a chance of *self-selection into a treatment*.

Unfortunately, it is difficult to resolve the above-discussed issues using suitable parametric approaches such as OLS estimates and Instrumental Variable (IV) estimators because the first one imposes a conditional independence assumption. In contrast, it is difficult to identify a relevant and exogenous instrumental variable in the second case. Moreover, both the IV and OLS approaches enforce a linear functional form assumption, which is arbitrarily ad hoc in that those coefficients on control variables are restricted to be the same for adopters and non-adopters. Therefore, the study deal with nonparametric methods to overcome the restrictive assumptions.

2.5.3 The propensity score matching procedure

Adopting any new technology (here mobile-app) is a function of various observable attributes at the household level, and eliminating the assumption of *constant technology impact* permits us to follow the PSM method. This approach balances the distributions of observed baseline covariates between the control and treatment groups depending on their predicted probabilities of adopting advanced technology like mobile-app (their '*propensity score*').

The major advantage of the treatment approach is that it creates a condition like a *randomized experiment* to estimate the causal impact, as in the case of a controlled experiment. To perform PSM, we need to follow the *conditional independence assumption*, which dictates that mobile-app adoption is random and not correlated with paddy yield once we control for X. Hence, we can express the mobile-app adoption effect as

$$\omega(X) = E(Y^1 - Y^0 \mid X) = E(Y^1 \mid M = 1, X) - E(Y^0 \mid M = 0, X) \tag{13}$$

Where the average effect of mobile-app is

$$\omega = E\{\omega(X)\}. \tag{14}$$

Since mobile-app adoption is random, we can compare paddy productivity of similar households with different socioeconomic statuses (i.e., either adopters or non-adopters) based on the values of X_s (baseline covariates). However, the baseline covariates differ widely, and the PSM procedure reduces this covariate dimensionality by comparing households with the same probability of adopting the mobile app, given the relevant controls X (Rosenbaum & Rubin, 1983).

Therefore, the study wants to identify the conditional probability that household i adopts the new mobile app, given the controls X as

$$\rho_i = \rho(X_i) = \mathrm{Pr}ob[M_i = 1 \mid X_i] \tag{15}$$

This conditional probability, known as propensity score, helps us identify similar households (Rosenbaum & Rubin, 1983). The propensity score ranks households based on their behavior towards mobile app adoption to estimate mobile-app impact among households with similar behavior. In other words, groups of farmers with the same propensity score have the same distribution of X, irrespective of their mobile-app adoption. It is termed a *balancing property*. Estimating balancing property is crucial to check whether farmers' behavior within each group is similar or not.

The mobile-app adoption impact for a group of farmers with similar propensity scores can be rewritten as

$$\omega(\rho(X)) = E\big(Y^1 \mid M = 1, \rho(X)\big) - E\big(Y^0 \mid M = 0, \rho(X)\big) \tag{16}$$

Where the effect on the entire population is

$$\omega = E\{\omega(\rho(X))\}. \tag{17}$$

The propensity score also imposes *common support* conditions as its value stays between 0 and 1. Such a value distribution pattern improves the matching quality as it excludes the tails of the distribution of $\rho(X)$. Hence, the PSM method is only meaningful as it applies to the area of overlapping support (Heckman et al., 1997).

After identifying the similarities through propensity score estimation, we match each adopter with their *nearest* non-adopter based on the similarity score. There are various methods to perform this activity. One such method is the *nearest neighbor* method (NNM) which identifies the *closest twin* for each household in the opposite adoption status. Then it determines the average difference in yield between each pair of matched households and concludes that the difference in yield arises due to mobile-app adoption. The second approach is known as a *kernel-based matching* estimator (KBM). This method is more flexible compared to NNM. It follows the same procedure as NNM, but the matched household is the weighted average of all households in the opposite adoption status within a certain propensity score distance, while weights are inversely proportional to the distance.

3. Results and Discussion

Different paddy inputs such as land rent, human labour, land area, machinery, and fertilizing chemicals, while production yield and CO2eq of GHG emissions as good and bad outputs, respectively, were included in the dataset. Irrigation cost has not been considered among inputs since paddy cultivation in this region is mostly rainfed. Descriptive statistics of different input and output variables have been summarized in Table 1 below. It can be observed that most of the farmers in this region are small landholders for whom paddy is the major source of income. Among inputs, renting farm machinery like tractors and harvesters is the farmers' major cost, whereas labor (both man and woman) is abundantly available. However, the yield is not very high compared to the state or national average. On the other hand, there are significant GHG emissions from the various stages of paddy production, an estimate of which has been calculated based on the yield in line with other research studies (Baruah et al., 2004; Chauhan et al., 2006; Karstensen et al., 2020).

(Insert Table 1 here)

3.1 DEA Efficiency Scores

DEA efficiency scores were evaluated both with and without considering undesirable outputs for comparison. Results show that the TE of 119 out of 153 farmers decreased after including GHG emissions as an undesirable output. Only the remaining 34 farmers had been able to manage their farm inputs in a better way and produced comparatively lower levels of GHG emissions. Similarly, the total number of efficient DMUs without undesirable output was found to be 52. However, this reduced significantly to 34 after including undesirable outputs in the analysis. This shows that incorporating undesirable outputs does lower the efficiency of DMUs, and the efficiency scores thus obtained are closer to the actual TE of the DMUs. The details of efficiency scores with and without undesirable outputs have been given in Table 2 below.

(Insert Table 2 here)

3.2 Slacks Analysis

Slacks in DEA refer to the proportional reduction in the input variables or increase in the quantity of the output variables, which is required for an inefficient DMU to become fully efficient. Slacks can not only recognize the inefficiencies farmers face in improving their farm productivity, but they can also provide appropriate direction where the improvement is most needed. Therefore, slacks represent the remaining portion of the efficiency and exist only for inefficient DMUs (Tone, 2001). It should be noted that the slack values on both input and output variables have zero values when the efficiency score equals one. Calculating slacks in inputs and outputs for individual DMUs provides us with tangible information regarding the amount by which each input variable must be changed to enable any particular DMU to become efficient. For this study, the slack values have been calculated using the multi-stage method (Tone, 2001; Yang, 2014).

Table 3 above shows the general descriptive statistics of the slacks, which can be defined as the space of improvement by re-allocating the variables. In the case of the input-oriented DEA model, the slacks in input variables indicate the scope for reduction, while the slacks in output variables represent the scope for potential improvement. The table shows that the cost of women's agricultural labor could be reduced by ₹218 on average. Similarly, the cost of fertilizers, farming machinery, and labor (adult man) should be reduced by approximately ₹1259, ₹1862, and ₹929, respectively. However, there is little scope for reduction in the area under paddy cultivation, which is helpful since paddy is these farmers' primary livelihood source. If individual farmers optimize the input slacks correctly, there is a potential scope of improvement in the yield by nearly 12 quintals on average and a reduction of 1397 kg of CO_2eq released during cultivation. Theoretically, these recommendations can help farmers reduce their cost of production and increase their overall productivity.

(Insert Table 3 here)

3.3 Socioeconomic status of the paddy farmers

Table 4 displays the socioeconomic condition of the paddy growers by mobile-app adoption status for 152 surveyed households using land for paddy production during the Kharif season. Explanatory variables in table 4 are selected based on theoretical assumptions and focused group discussion and used as baseline covariates to calculate the propensity score.

We identified that the average education level of the households' head and farm experience is statistically different between non-adopters and adopters, which means that education might be correlated with the adoption choice. A significant difference in the average landing size exists between adopters and non-adopters, which signifies that farmers' adoption decisions may be influenced by farm size. Among the institutional assets such as membership in farmers' groups, having Kisan Credit Card (KCC), and access to credit differ significantly among non-adopters and adopters, suggesting that these variables might act as critical factors to access mobile-app. Moreover, non-adopters present a significantly lower percentage of market information and investment in advanced mechanization and a higher percentage in livestock ownership than mobile-app adopters. Hence, such variables were included in the study as baseline covariates to estimate propensity scores and identify their impact on farmers' adoption decisions. Most importantly, the mobile-app adopters achieve a 15.62% (0.52t/ha) better yield than non-adopters, which motivates us to rectify whether the such impact is only due to mobile-app adoption or socioeconomic and institutional constraints is responsible for such variation in production among adopters and non-adopters.

(Insert Table 4 here)

3.4 Determinants of mobile app adoption

Table 5 summarizes the outcome of probit estimation. The result shows that the household head’s education level is important in adopting mobile-app. One additional year in formal education from the mean education level improves the likelihood of mobile-app adoption by 1.7%. Educated farmers tend to have a better possibility to decode new information and analyze the importance of advanced technology like mobile-app. Education facilitates adoption and helps better manage input resources (Alene & Manyong, 2007). This finding is similar to Kassie et al. (2011) and Khonje et al. (2015).

Like education, experienced farmers are more prone to technology adoption. A practical, timely, and optimum amount of input is the key to successful farming in terms of economic evaluation. Experienced farmers know these facts, and as the mobile app provides this information effectively, they are more interested in using this mobile app.

Social capital is crucial for technology adoption (Abebaw & Haile, 2013; Alene et al., 2008; Fischer & Qaim, 2014). This is also the case in this study, as the farmers with group membership show a 19.6% higher probability of mobile-app adoption than those without group membership. Group members share their experiences and exchange information about their new adoption when they meet, and a positive experience about an adoption always motivates farmers to adopt the technology (Kassie et al., 2011).

The results also indicate that access to market information significantly impacts farmers’ mobile-app adoption. A 1% increase in market information improves the chance of mobile-app adoption by 40%. This suggests that if paddy growers have market information and easy market access, it will reduce the high transaction cost in the quest to find markets for the input purchase and produce sell, which in turn will help them in gaining maximum benefit from the adoption of advanced technologies (Khonje et al., 2015).

Having Kisan Credit Card affects mobile-app adoption. Abate et al. (2016) identified that access to formal credit improved the chance of technology adoption. In India, farmers with Kisan Credit cards can avail of institutional credit (at a meager interest rate) before every cropping season without visiting formal lending institutions. With access to credit, farmers have sufficient working capital to buy farm inputs. However, they don’t know enough about the efficient use of inputs to get maximum farm output. However, the newly introduced mobile app provides them the platform to use the inputs efficiently for maximum benefit. Therefore, having a KCC improves the likelihood of app adoption by 26.4%.

Successful adoption of modern technologies always motivates and gives confidence about future smart technology adoption. This may be why those farmers who invested and adopted any new technology in the past also adopted the mobile app. However, age always plays a role in any kind of adoption. For advanced technology adoption, the young generation is more passionate; hence relatively young farmers adopt the mobile app compared to the older ones. Also, old farmers are unfamiliar with smartphone use; therefore, they are more reluctant to use any mobile app.

Results indicate that household size negatively impacts mobile-app adoption, suggesting that subsistence pressure might play a critical role in choosing advanced technology. Marginal landholders with large household sizes are mostly involved in subsistence farming. Most of the time, they are least interested in farming; hence, they are less likely to invest in any new technology for production improvement in the farm sector. Similarly, farmers with livestock ownership feel that livestock farming is more profitable than crop production. Therefore, such farmers are not interested in adopting any technology related to conventional crop production.

Asset ownership positively influences technology adoption (Khonje et al., 2015). If farmers have more cultivable land, they are more interested in farming because, with comparatively large landholding, they are capable of surplus production, and by the market this surplus amount, they gain profit. Such business-minded farmers always accept any new technology that increases their chance of profit. Perhaps this is why farmers with large landholding sizes are more interested in mobile-app adoption than others with comparatively less landholding sizes.

(Insert Table 5 here)

3.5 The causal effect of mobile-app adoption on poverty reduction

Before discussing the causal effects of mobile-app adoption on paddy yield, we need to identify the quality of the matching process. The initial requirement is to balance the distribution of relevant variables between mobile adopters and non-adopters. Table 6 provides the detailed results of the covariate balancing test before and after the matching. The standardized mean difference for overall covariates used in PSM is reduced by 56% to 60%, irrespective of the matching algorithm. This finding indicates a substantial reduction of total bias through matching. Also, the ρ-values of the likelihood ratio tests specify that the joint significance of covariates is not rejected before matching. However, it is rejected after matching. Also, the pseudo-R2 values for all three matching algorithms drop significantly after matching. The insignificant ρ-values of the likelihood ratio test, low pseudo-R2, and low mean standardized bias indicate that the propensity score successfully balances the distribution of covariates between the two groups.

(Insert Table 6 here)

The PSM estimates (NNM, KNM, and Radius Matching) presented in Table 7 show that farmers who adopted mobile-app had increased their paddy productivity. Mobile-app users achieve an additional 0.6t/ha paddy yield, which signifies that the mobile app helps the farmer provide optimum inputs (Fertilizer, irrigation, insecticides, pesticides, etc.) at the right time. An optimum and timely supply of inputs helps improve plant nutrition and growth, resulting in greater production per unit area.

The results obtained in Table 7 depend on the postulation of conditional independence and confoundedness. If any unobserved independent variable is present that can affect both mobile-app adoption and outcome variables, then the chance of unobserved heterogeneity appears, which can alter the influence significance (Becker & Ichino, 2002; Rosenbaum & Rubin, 1983). In non-experimental studies, it is difficult to determine the magnitude of such hidden bias due to the unavailability of a relevant measurement tool. Rosenbaum gave one feasible solution in 2002. Since then, by calculating the Rosenbaum bounds sensitivity analysis, we can determine how strongly the unobserved exogenous variables influence the significance of the estimate (Caliendo & Kopeinig, 2008; DiPrete & Gangl, 2004).

Results in Table 7 show that a τ-bound value is associated with each ATT value. Such value displays a critical gamma level at which the causal inference of mobile-app adoption may be questioned. For example, the gamma value for paddy is 2.90-2.95, which means that if households have the same vectors of baseline covariates in their odds of mobile-app adoption by a factor of 190-195%, the positive impact of mobile-app adoption on paddy yield, may be questionable. It means that the strength of hidden bias must be high enough to alter the findings in Table 7. It also indicates that the study has considered almost all those possible exogenous variables as baseline covariates that impact treatment and dependent variables.

(Insert Table 7 here)

4. Conclusions

Paddy is a resource-intensive crop. Optimum utilization of inputs is necessary to improve paddy yield, farmers' income, and greenhouse gas emissions. Hence, the objectives of this study were to identify the resource-inefficient farmers through DEA and provide them with mobile-based recommendations for the optimum use of farm inputs. Finally, the study also identified the impact of mobile app-based input recommendation adoption on farmers' yield improvement. The findings from DEA indicated that only 22% of farmers manage their farm inputs optimally. Slack analysis indicated that the cost of fertilizers, farming machinery, women's agricultural labor, and man-labor should be reduced by approximately ₹1259, ₹1862, ₹218, and ₹929, respectively. Based on the findings, we recommended inputs among the farmers with a mobile app to improve their farm production. Finally, we collected data from app users and non-users and performed a quasi-experimental analysis to identify whether the app-based input recommendation improved farmers' paddy yield. Results estimated that app users achieved an additional 0.6t/ha paddy yield than non-app users. We also determined the constraints for mobile app adoption. The findings indicated that education level, farm experience, social capital, market information, credit access, and prior use of advanced technology played a significant role in mobile app adoption. Such findings indicated that the government needs to create more digital platforms to provide farm-based advisories to smallholder farmers that will sustainably improve farm productivity. Along with the government, non-government organizations and private players can demonstrate digital platform utilization's beneficial role in enhancing their acceptability among farmers.

References

Abate, G. T., Rashid, S., Borzaga, C., & Getnet, K. (2016). Rural Finance and Agricultural Technology Adoption in Ethiopia: Does the Institutional Design of Lending Organizations Matter? *World Development*, *84*, 235–253. https://doi.org/10.1016/j.worlddev.2016.03.003

Abebaw, D., & Haile, M. G. (2013). The impact of cooperatives on agricultural technology adoption: Empirical evidence from Ethiopia. *Food Policy*, *38*(1), 82–91. https://doi.org/10.1016/J.FOODPOL.2012.10.003

Alene, A. D., & Manyong, V. M. (2007). The effects of education on agricultural productivity under traditional and improved technology in northern Nigeria: an endogenous switching regression analysis. *Empirical Economics*, *32*(1), 141–159.

https://doi.org/10.1007/s00181-006-0076-3

Alene, A. D., Manyong, V. M., Omanya, G., Mignouna, H. D., Bokanga, M., & Odhiambo, G. (2008). Smallholder market participation under transactions costs: Maize supply and fertilizer demand in Kenya. *Food Policy*, *33*(4), 318–328. https://doi.org/10.1016/J.FOODPOL.2007.12.001

Amid, S., Mesri Gundoshmian, T., Shahgoli, G., & Rafiee, S. (2016). Energy use pattern and optimization of energy required for broiler production using data envelopment analysis. *Information Processing in Agriculture*, *3*(2), 83–91. https://doi.org/10.1016/J.INPA.2016.03.003

Banaeian, N., Omid, M., & Ahmadi, H. (2012). Greenhouse strawberry production in Iran, efficient or inefficient in energy. *Energy Efficiency*, *5*(2), 201–209. https://doi.org/10.1007/S12053-011-9133-7/FIGURES/2

Banaeian, N., & Zangeneh, M. (2011). Study on energy efficiency in corn production of Iran. *Energy*, *36*(8), 5394–5402. https://doi.org/10.1016/J.ENERGY.2011.06.052

Barra, C., & Zotti, R. (2016). A Directional Distance Approach Applied to Higher Education: An Analysis of Teaching-Related Output Efficiency. Annals of Public and Cooperative Economics, 87(2), 145–173. https://doi.org/10.1111/apce.12091

Baruah, D. C., Das, P. K., & Dutta, P. K. (2004). Present status and future demand for energy for bullock-operated paddy-farms in Assam (India). Applied Energy, 79(2), 145–157. https://doi.org/10.1016/j.apenergy.2003.12.014

Becker, S. O., & Ichino, A. (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *The Stata Journal: Promoting Communications on Statistics and Stata*, *2*(4), 358–377. https://doi.org/10.1177/1536867X0200200403

Caliendo, M., & Kopeinig, S. (2008). SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/J.1467-6419.2007.00527.X

Carlsson-Kanyama, A., & González, A. D. (2009). Potential contributions of food consumption patterns to climate change. *The American Journal of Clinical Nutrition*, *89*(5). https://doi.org/10.3945/AJCN.2009.26736AA

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimization Theory and Applications*, 98(2), 351–364. https://doi.org/10.1023/A:1022637501082

Chandio, A. A., Jiang, Y., Rehman, A., Twumasi, M. A., Pathan, A. G., & Mohsin, M. (2020). Determinants of demand for credit by smallholder farmers': a farm level analysis based on survey in Sindh, Pakistan. *Journal of Asian Business and Economic Studies*, *28*(3), 225–240. https://doi.org/10.1108/JABES-01-2020-0004

Chauhan, N. S., Mohapatra, P. K. J., & Pandey, K. P. (2006). Improving energy productivity in paddy production through benchmarking—An application of data envelopment analysis. *Energy Conversion and Management*, *47*(9–10), 1063–1085. https://doi.org/10.1016/J.ENCONMAN.2005.07.004

DiPrete, T. A., & Gangl, M. (2004). 7. Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, *34*(1), 271–310.

https://doi.org/10.1111/j.0081-1750.2004.00154.x

Emrouznejad, A., Podinovski, V. v., & Thanassoulis, E. (2009). Data envelopment analysis: Theory and applications. *Journal of the Operational Research Society*, 60(11), 1467–1468. https://doi.org/10.1057/jors.2009.73

Erdal, G., Esengün, K., Erdal, H., & Gündüz, O. (2007). Energy use and economical analysis of sugar beet production in Tokat province of Turkey. *Energy*, *32*(1), 35–41. https://doi.org/10.1016/J.ENERGY.2006.01.007

FAO. (2012). The State of Food and Agriculture. Food and Agriculture Organization of the United Nations, Rome, 2012.

FAO. (2004). The State of Food Insecurity in the World. Food and Agriculture Organization of the United Nations, Rome, Italy.

Fischer, E., & Qaim, M. (2014). Smallholder Farmers and Collective Action: What Determines the Intensity of Participation? *Journal of Agricultural Economics*, *65*(3), 683–702. https://doi.org/10.1111/1477-9552.12060

Flückiger, Y., & Vassiliev, A. (2007). Efficiency in Microfinance Institutions: An Application of Data Envelopment Analysis to MFIs in Peru. In B. Balkenhol (Ed.), Microfinance and Public Policy (1st ed., pp. 89–110). Palgrave Macmillan UK. https://doi.org/10.1057/9780230300026_6

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, *66*(5), 1017. https://doi.org/10.2307/2999630

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, *64*(4), 605–654. https://doi.org/10.2307/2971733

Hosseinzadeh-Bandbafha, H., Safarzadeh, D., Ahmadi, E., Nabavi-Pelesaraei, A., & Hosseinzadeh-Bandbafha, E. (2017). Applying data envelopment analysis to evaluation of energy efficiency and decreasing of greenhouse gas emissions of fattening farms. *Energy*, *120*, 652–662. https://doi.org/10.1016/J.ENERGY.2016.11.117

Houshyar, E., Azadi, H., Almassi, M., Sheikh Davoodi, M. J., & Witlox, F. (2012). Sustainable and efficient energy consumption of corn production in Southwest Iran: Combination of multi-fuzzy and DEA modeling. *Energy*, *44*(1), 672–681. https://doi.org/10.1016/J.ENERGY.2012.05.025

IRRI. (2013). CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Copenhagen, Denmark (2014) 25 pp.

Karstensen, J., Roy, J., Pal, D., Peters, G., & Andrew, R. (2020). Key Drivers of Indian Greenhouse Gas Emissions. *Economic and Political Weekly*, 55(15), 46–53. https://www.epw.in/journal/2020/15/special-articles/key-drivers-indian-greenhouse-gas-emissions.html?0=ip_login_no_cache%3D8ac02bb000c461db744be5fe2f69189b

Kassie, M., Shiferaw, B., & Muricho, G. (2011). Agricultural Technology, Crop Income, and Poverty Alleviation in Uganda. *World Development*, *39*(10), 1784–1795. https://doi.org/10.1016/J.WORLDDEV.2011.04.023

Khonje, M., Manda, J., Alene, A. D., & Kassie, M. (2015). Analysis of Adoption and Impacts of Improved Maize Varieties in Eastern Zambia. *World Development*, *66*, 695 706-706.

https://doi.org/10.1016/j.worlddev.2014.09.008

Kwon, H. B., & Lee, J. (2015). Two-stage production modeling of large U.S. banks: A DEA-neural network approach. *Expert Systems with Applications*, 42(19), 6758–6766. https://doi.org/10.1016/j.eswa.2015.04.062

Malana, N. M., & Malano, H. M. (2006). Benchmarking productive efficiency of selected wheat areas in Pakistan and India using data envelopment analysis. *Irrigation and Drainage*, 55(4), 383–394. https://doi.org/10.1002/ird.264

Mobtaker, H. G., Keyhani, A., Mohammadi, A., Rafiee, S., & Akram, A. (2010). Sensitivity analysis of energy inputs for barley production in Hamedan Province of Iran. *Agriculture, Ecosystems & Environment*, *137*(3–4), 367–372. https://doi.org/10.1016/J.AGEE.2010.03.011

Mousavi-Avval, S. H., Rafiee, S., Jafari, A., & Mohammadi, A. (2011). Optimization of energy consumption for soybean production using Data Envelopment Analysis (DEA) approach. *Applied Energy*, 88(11), 3765–3772. https://doi.org/10.1016/j.apenergy.2011.04.021

Muthayya, S., Sugimoto, J. D., Montgomery, S., & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences*, *1324*(1), 7–14. https://doi.org/10.1111/NYAS.12540

Nghiem, H. S., Coelli, T., & Rao, P. (2007). The Efficiency of Microfinance in Vietnam: Evidence from NGO Schemes in the North and the Central Regions. *The International Journal of Environmental, Cultural, Economic, and Social Sustainability: Annual Review*, 2(5), 71–78. https://doi.org/10.18848/1832-2077/CGP/v02i05/54258

Phadnis, S. S., & Kulshrestha, M. (2012). Evaluation of irrigation efficiencies for water users' associations in a major irrigation project in India by DEA. *Benchmarking: An International Journal*, 19(2), 193–218. https://doi.org/10.1108/14635771211224536

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLOS ONE*, *8*(6), e66428. https://doi.org/10.1371/JOURNAL.PONE.0066428

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Sengupta, J. K. (1987). Production frontier estimation to measure efficiency: A critical evaluation in light of data envelopment analysis. *Managerial and Decision Economics*, 8(2), 93–99. https://doi.org/10.1002/mde.4090080203

Soni, P., Sinha, R., & Perret, S. R. (2018). Energy use and efficiency in selected rice-based cropping systems of the Middle-Indo Gangetic Plains in India. *Energy Reports*, 4, 554–564. https://doi.org/10.1016/j.egyr.2018.09.001

Tiongco, M., & Dawe, D. (2002). Long-term Evolution of Productivity in a Sample of Philippine Rice Farms: Implications for Sustainability and Future Research. *World Development*, *30*(5), 891–898. https://doi.org/10.1016/S0305-750X(02)00011-6

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130(3), 498–509. https://doi.org/10.1016/S0377-2217(99)00407-5

Yang, C. C. (2014). Evaluating the performance of banking under risk regulations: A slacks-based Data Envelopment Analysis assessment framework. *Expert Systems*, 31(2), 176–184. https://doi.org/10.1111/exsy.12020

Yuan, S., & Peng, S. (2017). Input-output energy analysis of rice production in different crop management practices in central China. *Energy*, 141, 1124–1132. https://doi.org/10.1016/J.ENERGY.2017.10.007

Zhou, Z., Amowine, N., & Huang, D. (2018). Quantitative efficiency assessment based on the dynamic slack-based network data envelopment analysis for commercial banks in Ghana. *South African Journal of Economic and Management Sciences*, 21(1), 1–11. https://doi.org/10.4102/sajems.v21i1.1717

Tables & Figure

Table 1. Descriptive Statistics of different Inputs and Outputs

S. No.	Particulars	Unit	Mean	SD	Min	Max
Input Variables						
1	Seed cost	*INR/ha*	3842	378.23	2575	5380
2	Hired labour cost	*INR/ha*	4133.01	204.48	1575	7325
3	Family labour (Adult Man)	*Hours/ha*	169.28	1150.36	0	4800
4	Family labour (Adult Woman)	*Hours/ha*	72.29	520.18	0	3000
5	Electricity	*kWh/ha*	148.29	48.69	82.32	236.91
6	Diesel fuel	*Litre/ha*	80.88	50.76	28.15	151.72
7	Nitrogen fertilizer (N)	*(kg/ha)*	107.23	22.28	70.91	140.24
8	Phosphorus fertilizer (P_2O_5)	*(kg/ha)*	53.68	12.16	40.11	70.63
9	Potassium fertilizer (K_2O)	*(kg/ha)*	49.92	13.63	35.73	64.77
Output Variables						
1	Paddy production (desirable)	*Tons/ha*	2.82	2.29	1.73	4.73
2	Total CO_2eq (undesirable)	*kg of CO_2*	1396.91	789.43	332.28	4033.24

Table 2. Summary of DEA efficiency scores

S. No.	Particulars	Mean	Std. Dev.	Min	Max
1	Efficiency without undesirable o/p	0.814668	0.170925	0.41168	1
2	Efficiency with undesirable o/p	0.639601	0.242239	0.22288	1

Table 3. Summary of Slacks for DEA with undesirable output

S. No.	Particulars	Unit	Mean	SD	Min	Max
Input Variables						
1	Land Size	*Ha*	0.12	0.16	0	0.78
2	Land Rent	₹	1395.12	1447.56	0	6500
3	Labour (Adult Man)	₹ *per day*	928.65	997.23	0	3850
4	Labour (Adult Woman)	₹ *per day*	217.64	380.86	0	2500
5	Farming Machinery Cost	₹	1861.55	2299.53	0	9350.96
6	Cost of Fertilizers/ Pesticides	₹	1258.97	1884.17	0	10412.5
Output Variables						
1	Paddy Yield (desirable)	*Quintals*	11.76	6.29	3.73	31.73
2	Total CO_2eq (undesirable)	*kg of CO_2*	1396.91	789.43	332.28	4033.24

Table 4. Descriptive statistics for mobile-app adopters and non-adopters.

Variables	Mobile-app non-adopters		Mobile-app adopters		Mean difference test
	Mean	SE	Mean	SE	
Yield	3.33	0.025	3.85	0.04	0.001
Age	46.64	1.5	45.44	2.84	0.697
Gender	0.833	0.035	0.763	0.069	0.337
Education	6.79	0.302	9.18	0.609	0.002
Farm experience	31.89	1.7	34.11	2.85	0.515
Farm size	0.378	0.021	1.02	0.075	0.001
Household members	5.27	0.167	5.18	0.303	0.796
Farming as major income source	0.62	0.045	0.76	0.069	0.942
Distance to marketplace	2.65	0.095	2.32	0.166	0.083
Membership in farmers group	0.342	0.044	0.711	0.074	0.001
KCC	0.298	0.043	0.736	0.072	0.001
Market information	0.263	0.041	0.631	0.079	0.001
Access to credit	0.491	0.047	0.68	0.076	0.039
Livestock	0.711	0.042	0.447	0.081	0.003
Prior investment in advanced mechanization	0.272	0.041	0.447	0.081	0.044

Table 5. Probit estimates of the determinants of mobile-app.

Variables	**Estimates**	**SE**	**Marginal effect**
Age	-0.112***	0.043	-0.016
Gender	-0.359	0.484	-0.052
Education	0.119**	0.063	0.017
Farm experience	0.115***	0.042	0.016
Farm size	0.108*	0.042	0.012
Household members	-0.225**	0.109	-0.032
Farming as major income source	-0.078	0.389	-0.011
Distance to marketplace	-0.213	0.193	-0.031
Membership in farmers group	1.346***	0.518	0.196
KCC	1.817***	0.562	0.264
Market information	2.745***	0.615	0.4
Access to credit	0.342	0.422	0.049
Livestock	-0.817**	0.437	-0.119
Prior investment in advanced mechanization	0.654*	0.39	0.095
Constant	-0.872	1.165	

*** Significant at 1%; ** Significant at 5%; * Significant at 10%

Table 6. PSM quality indicators before and after matching

	Matching algorithm	Pseudo-R^2 before matching	Pseudo-R^2 after matching	$\rho>\chi 2$ before matching	$\rho>\chi 2$ after matching	Mean standardized bias before matching	Mean standardized bias after matching	(Total)% |bias| reduction
Mobile-app adoption	NNM	0.483	0.083	0.007	0.538	0.353	0.143	59.49
	KBM	0.476	0.073	0.009	0.389	0.361	0.155	57.06
	RM	0.453	0.080	0.011	0.441	0.376	0.163	56.64

Table 7. Average treatment effects on treated and results of sensitivity analysis

Matching algorithm	**ATT**	**SE**	**τ-bound**
NNM	0.594***	0.171	2.60-2.65
KBM (0.06)	0.609***	0.039	2.20-2.25
RM (0.25)	0.574***	0.093	2.90-2.95

*** Significant at 1%

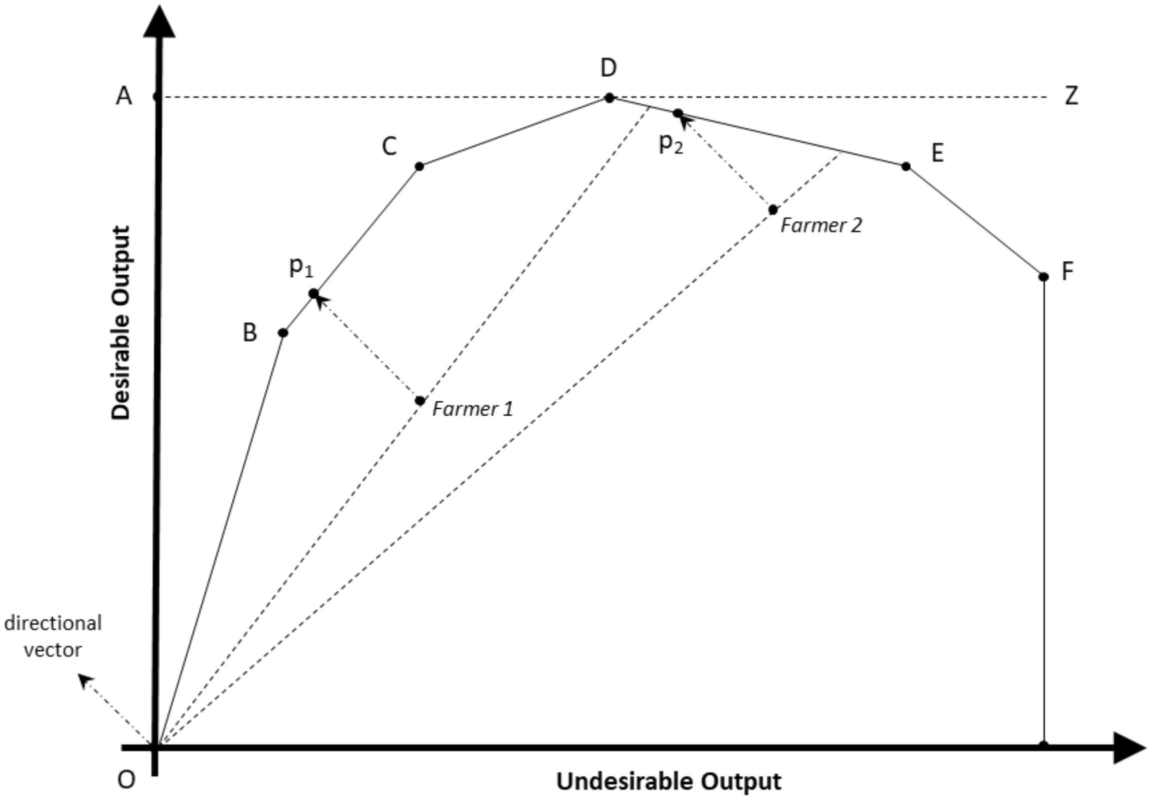

Figure 1. An environmental PPS under undesirable output-based DEA