5-2018

# An Impact Evaluation of the Government Grant Programs and Input Efficiency Use in the Production of Greenhouse Tomatoes and Peppers in Kosovo

Blend Frangu
*University of Arkansas, Fayetteville*

An Impact Evaluation of the Government Grant Programs and Input Efficiency Use in the Production of Greenhouse Tomatoes and Peppers in Kosovo

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Agricultural Economics

by

Blend Frangu
University of Prishtina "Hasan Prishtina"
Bachelor of Agro-economy, 2015

May 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Jennie S. Popp, Ph.D.
Thesis Director

Arben Musliu, Ph.D.
Committee Member

Michael R. Thomsen, Ph.D.
Committee Member

**Abstract**

Determinants of greenhouse tomato and pepper production input efficiency affect a farmer's decision to contribute to the vegetable sector in Kosovo. This research investigates the non-optimal use of inputs in the production of greenhouse tomatoes and peppers. Two studies were conducted to first measure input efficiency use and then to quantify the impact of the Ministry of Agriculture, Forestry and Rural Development (MAFRD) grant programs in the production of both crops. The first study applied data envelopment analysis (DEA), and linear and logistic regression to assess factors that influence efficiency in production. The second study used propensity score matching with logistic regression and genetic matching in order to evaluate any impacts, including income differences, between MAFRD grantees and non-grantees. The DEA results suggest that of the seven regions in Kosovo, most of the efficient greenhouse tomato producers were found in Prizren with a mean efficiency of 83 percent. While Prishtina had the most efficient greenhouse pepper producers with a mean efficiency of 99 percent. The logistic regression results with the use of technical efficiency (TE) scores as the dependent variable with a threshold produced different results to the linear regression. For example, estimates were more statistically significant under linear regression. The second study using matching techniques demonstrated that greenhouse farmers with higher yields and small total greenhouse areas were more likely to participate in the government's grant program. The analysis revealed that MAFRD greenhouse tomato grantees compared to the non-grantees may make additionally an estimated of 1,777 euros. Insignificant estimate results were registered for the sample of greenhouse pepper farmers. Overall, the first study demonstrated that depending on the characteristics of the farms, factors related to production efficiency may affect input efficiency use. The second study suggested that a quality matching of the greenhouse tomato farmers could be achieved. However, only a partial matching was obtained among greenhouse pepper farmers.

**Table of Contents**

## Chapter 1. Introduction

### Greenhouse Production

Kosovo in Southeastern Europe was part of Yugoslavia with an underdeveloped social structure and low level of economic development (Elizabeth, 1981). The dynamics of Kosovo's economy were shaped by the Kosovo war (1998-1999), a conflict that left the region's agriculture vulnerable and with a loss of productivity. Later, Kosovo declared its independence in 2008 and became a young country with the objective of building social and economic institutions (International Monetary Fund, 2011).

Over the last decade, the European Union (EU) block has remained the largest trading partner to Kosovo (Gashi, 2017). Yet, the high level of imports from Macedonia, Albania and Turkey have undermined Kosovo's agricultural base (Archer, 2003), and the excessive volume of imports has generated a negative trade balance (Jusufi, et al., 2015). The issue of year-round imports grabbing farmers' sales in the market has grown over time. During the vegetable season, competition posed by vegetable crop imports complicates domestic farmers' ability to sell in the market. This challenge requires changes in the way vegetable crops are produced internally.

Agriculture is Kosovo's initial source of economic development. The new presence of greenhouse tomato and pepper farms marks an emerging subsector. Farmers in this subsector face land fragmentation, high imports, and low productivity levels; thus, they understand the difficulties involved in selling in the market and may decide to use inputs in a way that reduces costs. Minimizing costs in the production of greenhouse tomatoes and peppers often happens so that farmers may maintain the overall financial health of the farm. Farmers dealing with low productivity levels raise the question of how to use inputs more efficiently. There is confusion among Kosovar producers over the possible ways to achieve greenhouse tomato and pepper input efficiency use. One common and not yet proven way suggests that government agricultural policies may provide incentives and guidelines to improve input efficiency. Studies may help provide an understanding of how to maintain current yields while decreasing waste of inputs as well as how to quantify the impact and effectiveness of government agricultural policies in production and in input efficiency use.

**Government's Role**

Foreign agencies for development, government institutions, and nonprofit organizations play a leading role in promoting efficient cultivation of the greenhouse vegetable crops in Kosovo. Often, optimal greenhouse production may require a transition from the traditional single tunnel greenhouses to multi tunnel greenhouses. In fact, greenhouses with improved design help farmers gain a competitive edge in the market due to the early season production (Balliu & Kaçiu, 2008). However, more general work to encourage the transition from traditional greenhouses to greenhouses with improved designs and upgraded farm facilities further strengthens the production levels of this fragile yet growing subsector of the agriculture industry.

To promote vegetable production, Kosovo's Ministry of Agriculture, Forestry and Rural Development or MAFRD (2016) initiated one of its funding activities titled "Measure 101: Investments in Physical Assets and Agricultural Households". In 2015, this investment activity funded 65 (or 13.6 percent) of 479 applications received from farmers to acquire new greenhouses, support for open-field production and/or storage warehouses. These serve as farm facilities which are important in the production of vegetable crops and reached a value of 3,275,340.85 euros in 2015 (MAFRD, 2016). The high application rates suggest that farmers have a great need for on-farm facility upgrades.

This governmental financial support has helped some farmers purchase new greenhouses with improved design and equipment. Farmers who have not received financial support argued that high-priced greenhouses prevent them from growing year-round. Another issue of importance is the low levels of productivity in the production of tomatoes and peppers in traditional and low tunnel greenhouses (Balliu & Kaçiu, 2008). In a recent study, Kaciu, Babaj, Aliu, and Demaj (2016) argue that more work is necessary to improve greenhouse production productivity and alleviate poor crop management practices. To emphasize the changes of this subsector, MAFRD (2016) showed that there were 14,656 planted vegetable hectares (ha) in 2015 relative to 15,854 ha in 2014. These marked increases of 21 percent and 42 percent of pepper and tomato areas in ha, respectively. The total pepper yield in 2015 had a value of 24,333 tons and 55,469 tons for tomatoes. Relative to 2014, total tomato output increased by 40 percent while total pepper output decreased by 4 percent. Overall, vegetable crop production decreased by 1198 ha or roughly 8 percent compared to 2014. Given these changes, it is necessary for the government to revisit the understanding of growing vegetable crops in Kosovo.

Under current circumstances, MAFRD (2017) with their detailed publication titled "Agriculture and Rural Development Program 2017" explains that the greenhouse vegetable sector has a limited number of commercial growers. According to this publication, greenhouse production is more labor intensive compared to open field production and with a real prospect for employment generation. It is important to note that labor plays a vital role in the development of the greenhouse production sector given the intensity of the work needed to perform the required farm operations. Graeub, et al. (2016) stated that a farm with less than 10 ha represents a farm with family labor as its primary source of workforce. Building on this line of research, Fall and Magnac (2004) stated that on-farm labor hours fall with the education level and raise with the number of adults and children in the farm. More research is necessary to explore variables that impact farmers' production efficiency, and whether research of this nature may improve the chances for an increasing number of competitive farmers in Kosovo.

This thesis contains two studies that aim to examine the state of the greenhouse tomato and pepper input efficiency use under the effect of MAFRD agricultural policies and under farmers' production management. The first study uses an input-oriented data envelopment analysis (DEA) coupled with the linear and binary logistic regression to quantify input efficiency use. The use of DEA allows a comparison of the performance of any given producer relative to the performance of the most efficient producer (Charles & Kumar, 2012) within our sample of greenhouse vegetable farms. Kosovar farmers rely heavily on efficiency improvements to expand their farm operations. The over or under utilization was an issue that this study's participants raised during the interview process. The farmers' desire was to understand what the optimal values of inputs are in the production process. Likewise, the objective of linear and binary logistic regression as a second step to DEA is to show us which additional variables have an impact in explaining input efficiency use.

The second study considers the practical aspects of government agricultural policies that have affected many farmers in seven regions of Kosovo: Prishtina, Ferizaj, Gjilan, Prizren, Gjakova, Peja, and Mitrovica. Farmers benefit from MAFRD's policies offering grants to purchase new greenhouses and equipment. The government grant programs and policies in agriculture have increased and yet few studies have looked at their effectiveness. Following the early counterfactual framework (Rubin, 1977; Rosenbaum & Rubin, 1983), the second study applies propensity score matching (PSM) with logistic regression as the distance measure and genetic matching to estimate the casual treatment effects and the seasonal income differences among MAFRD farmer grantees and non-grantees.

## First Study Methods Background

### *Data Envelopment Analysis*

In a commonly discussed study, Farrell (1957) measured efficiency by calculating a score for each observed firm and showed through an illustration how a pertinent estimate of the production function is obtained and applied to agricultural production in the United States. The extension and statistical formulation of this line of research was later developed by others (Aigner & Chu, 1968; Aigner, Lovell, & Schmidt, 1977; Meeusen & Van Den Broeck, 1977). Charnes, Cooper, and Rhodes (1978) extended Farrell's work with their elaboration of the concept of a decision making unit (DMU). The same authors (1978) used the term data envelopment analysis (DEA) to specify an efficiency measure in their research report. These findings allow the first study to use an input-oriented version of DEA as an efficiency measurement in the context of greenhouse tomato and pepper production.

A notable feature of DEA is that in contrast to the parametric methods such as stochastic frontier analysis (SFA), it does not assume that the data follows a specific distribution. This led to the understanding that the parametric approaches assume a specific form of the production function (Fried, Lovell, & Schmidt, 2008), while the same does not hold true for the non-parametric approaches (Johnes, 2006). A non-parametric approach may provide a relevant measure to quantify input efficiency use in the first study given the nature of the collected agricultural field data with a roughly normal distribution. Another distinction is that DEA is less susceptible than SFA to the specification error and with the advantage of considering multiple inputs and outputs (Reinhard, Lovell, & Thijssen, 2000). The inclusion of multiple inputs allows the first study of this thesis to evaluate the impact of numerous farm variables on greenhouse tomato and pepper yields.

*BCC and CCR Models*

The Charnes, Cooper, and Rhodes (CCR) (1978) DEA model with constant returns to scale (CRS) and the Banker, Charnes, and Cooper (BCC) model (1984) with variable returns to scale (VRS) made them become popular models for quantifying efficiency. The CCR model for each DMU forms the virtual input and output (yet with unknown values) by weights. Linear programming is used to determine and maximize the ratio of the specifically named virtual output to input. The weights derived from the data are different for each DMU, which are then compared to the relative efficiency ratios of the data. A DMU with a ratio value of 1.00 and zero-slack indicates CCR-efficiency, while an efficiency score below 1.00 indicates CCR-inefficiency (Cooper, Seiford, & Tone, 2002). The BCC model provides an efficiency score for a DMU that is not smaller than the CCR.

It follows that BCC deals with pure technical efficiency (PTE) under VRS and CCR deals with technical efficiency (TE) under CRS. The two models are different with respect to their assumptions on returns to scale. With CRS, each application of an extra unit of input produces the same amount of output. When in fact, VRS allows the extra output produced by a unit of extra input to vary in accordance with scale size. Under usual conditions, the CCR graphically exhibits an upward sloping line, while the BCC uses a piecewise linear curve to form the efficient frontier (Korhonen & Joro, 2015).

The direction of the models can be input or output-oriented depending on the objective of the researcher. The first study employs the input-oriented BCC and CCR models. The joint CCR- and BCC-efficiency analysis explains whether a greenhouse farm is operating under CRS or VRS. In other words, it shows whether a typical greenhouse tomato or pepper farm is over or underutilizing its inputs.

*Scale Efficiency*

The evaluation of DEA efficiency results often leads to a closely related notion, that of scale efficiency (SE). This is a useful concept to determine whether farms are operating optimally. In this case, SE shows correctly if a DMU is operating at its optimal size. Scale efficiency is a ratio of the CCR / BCC models and takes a value ranging from zero to one (Cooper, Seiford, & Tone, 2002). In the first study's frame of reference, greenhouse farms can be thought of as DMUs when used in the input-oriented DEA. If a greenhouse farm receives an SE score of one, it reveals that any further changes to its size may not lead to higher efficiency.

Scale efficiency can be analyzed through observing the efficient frontier. The increased SE maintains that a DMU has progressed to an improved position when its input-output ratio is achieved on the frontier (Balk, 2001). For example, if a greenhouse farm has performed no technical change and is on the efficient frontier, then this indicates only an SE movement. Similarly, Kumar and Gulati (2008) stated that the measure of SE does not show if a DMU is operating in the region of increasing returns to scale (IRS) or decreasing returns to scale (DRS).

*Regression Analysis*

Linear and binary logistic regression are models with wide applicability across a range of economic subdisciplines. The logistic model was developed by Cox (1958) and has a categorical dependent variable which takes a value of zero or one (Hayes & Matthes, 2009), and linear regression first published by Legendre (1805) has a scalar dependent variable. In contrast to linear regression, binary logistic regression has no formula for the beta estimates. The attempt to find the best beta estimates requires repeated improvement of the respective estimates until model stability is achieved (LaValley, 2008). Tolles and Meurer (2016) indicated that the logistic model has the capability to reveal factors that have the strongest impact on the outcome. For this purpose, the flexibility of the binary logistic regression is in its ability to account for confounding factors and Lever, Krzywinski, and Altman (2016) added to this saying that it has a relatively uncomplicated computation and interpretation.

For the first study, it is important to understand that binary logistic regression models the probability of an (in)efficient greenhouse as a response of its covariates, while linear regression determines the relationship of farm variables on TE scores. The CCR efficiency for the binary logistic model is the categorical dependent variable, where one shows an efficient greenhouse farm and zero an inefficient greenhouse farm. A farm identified as CCR efficient in the CCR model received a value of one and zero otherwise. Coding a binary dependent variable based on the received efficiency score value of one and as zero otherwise is discussed and showed by Tanfani and Testi (2012). However, in the linear model, CCR efficiency is used without a threshold and as a continuous variable for result comparison purposes.

Bursac, Gauss, Williams, and Hosmer (2008) established that an effective way to select variables for the model is the human modeling process. To identify in the first study the optimal set of variables for the regression models, an exhaustive search algorithm is performed in R programming coupled with human logic under Kosovo's context. The quality of the binary logistic model is found using the Akaike information criterion (AIC), and stability of the linear model was achieved using the adjusted R-squared.

## Second Study Methods Background

### *Propensity Score Matching with Logistic Regression*

The early work to develop propensity score matching (PSM) was initiated by Rosenbaum and Rubin (1983), and now has become a widely used approach to estimate causal treatment effects (Caliendo & Kopeinig, 2008). A remarkable feature of PSM is its ability to reduce the selection bias of the group receiving the treatment relative to the control group. The reduction of the selection bias occurs after PSM controls for covariates related to the treatment group. However, it may be less capable to control for unobserved selection bias (Guo & Fraser, 2014).

It is of interest in this study to conduct a complete analysis using logistic regression as an estimation procedure for PSM and genetic matching to examine the average treatment effect on the treated (ATT) on the greenhouse tomato and pepper grantees of the Ministry of Agriculture, Forestry and Rural Development (MAFRD). It follows logically that the estimation is dependent on the choice of the covariates. For the evaluation of the model, age, yield in kilogram (kg), greenhouse value in euros, greenhouse area in square meters ($m^2$), distance to the market in kilometers (km), and other greenhouse crops grown are included as covariates. Stuart (2010) explains that ATT represents the effect for subjects in the treatment group. As discussed by Abdia et al. (2017), detailed attention is paid to avoid a misspecification of the propensity score model, as it may provide biased estimates for ATT.

The estimation of propensity scores may be performed using binary logistic regression as a propensity score estimation method (McCaffrey, Ridgeway, & Morral, 2004). This method is consistent notably with observational data (Westreich, Lessler, & Funk, 2010), and its strength lies in its ability to include many variables (Hosmer, Lemeshow, & Sturdivant, 2013). Given the binary treatment case in the logistic model, this method estimates the probability of a subject’s participation versus nonparticipation (Caliendo & Kopeinig, 2008) with our sample of greenhouse tomato and pepper farmers.

To acquire the propensity scores of the farmers subject to propensity score matching (PSM) analysis, binary logistic regression is used. The potential of the model is to quantify the impact of widely different covariates on the binary outcome. In the view of many researchers, this method in agriculture was performed with success to assess factors that impact the adoption of vegetable crop diversification (Ali, 2013), natural forest harvesting of tobacco (Chivuraise, Chamboko, & Chagwiza, 2016), and tomato marketability (Tolesa, Workneh, & Melesse, 2017).

*Genetic Matching*

Propensity score matching can be performed using various methods to match subjects. One recent method includes genetic matching developed by Mebane and Sekhon (1998) as a multivariate matching method. With the genetic matching method, an evolutionary search algorithm is performed to determine the weight that each covariate of interest has been given in the specified model (Diamond & Sekhon, 2013). An additional feature of this matching algorithm is the choice of performing matching of the subjects with replacement or without replacement (Caliendo & Kopeinig, 2008).

In this study's context, matching is performed with replacement given the not too large number of observations and the small number of grantees (treatment group) compared to the non-grantees (control group). There is an advantage using this form of matching if greenhouse pepper relative to the greenhouse tomato farmer grantees and non-grantees show no large differences in covariate values. Stuart and Rubin (2008) suggest matching with replacement provides better matches. On this line of reasoning, Dehejia and Wahba (2002) add to this saying that matching with replacement has the potential to minimize the distance of the propensity scores between the control and treatment subjects.

**Kosovo's Context**

The presence of using data envelopment analysis (DEA) together with linear and binary logistic regression models is not widely used in the discipline of agriculture in Kosovo. Using these methods to conduct research in the agriculture sector may provide a possibility of what farm production-related challenges may be addressed in Kosovo. Depending on the subsector of agriculture subject to analysis and the objective of the researcher, variations of these models may be applied.

One extensive literature review did not identify studies using DEA and any method as a second step to DEA except Vuciterna (2017). The author measured the efficiency of Kosovo raspberry producers through employing input-oriented DEA and stochastic frontier analysis (SFA) models with the objective of providing more information for the competitiveness of the raspberry industry. However, studies using propensity score matching (PSM) may have a more significant research presence in Kosovo (see e.g. Sauer, Gorton, & Davidova, 2015; Bajrami, 2016).

No study has been identified in the extensive literature review using any of the models to evaluate the effectiveness of the Kosovar government agricultural policies on the greenhouse tomato and pepper productivity and of the government grant program effects on farmers' seasonal income. There is overall limited use of the models in farm production studies. Nevertheless, the increasing complexity of farming in Kosovo may result in further uses of input- or output-oriented DEA, PSM and linear and binary logistic regression to help identify policies that can lead to the overall efficiency of the agricultural sector in Kosovo.

## Chapter 2. Evaluating Greenhouse Tomato and Pepper Input Efficiency Use in Kosovo


### Abstract

Greenhouse tomato and pepper farms in Kosovo are constantly aiming to improve input efficiency use with the goal of increasing the gross margins. This study evaluates how efficiently a sample of greenhouse tomato and pepper farms use inputs in the production process to produce high yields. Using collected agricultural data, this study develops an input-oriented data envelopment analysis (DEA) model to empirically research input efficiency use. Secondly, a two-step analysis is developed through multiple linear and binary logistic regression analyses to investigate which farm variables predict greenhouse tomato and pepper technical efficiency (TE). The DEA results indicated that among the seven regions in Kosovo, Prizren emerged as the region with the most efficient greenhouse tomato producers with a mean efficiency of 83 percent. The region of Prishtina followed with a mean efficiency of 80 percent. While, in the production of greenhouse peppers, Prishtina had the most efficient producers with a mean efficiency of 99 percent. Ferizaj followed with a mean efficiency of 93 percent. The use of TE scores with a threshold to indicate an efficient greenhouse in the logistic regression model produced comparatively different results to using TE scores as the scalar dependent variable in the linear regression model. Depending on the structural and operational characteristics of the greenhouse tomato and pepper farms, overall different factors can affect input efficiency use.

## Introduction

Kosovo's fragile agriculture has been dependent on the performance of the agricultural machinery industry, and food and processing industries. The farmers of Kosovo went through many crises and conflicts. The 1998-1999 war left it with a collapse of the rural infrastructure, law and order (Judah, 2008), which impacted its vulnerable agriculture causing a loss of productivity. The Kosovar farmers' relationship with the land, however, has remained strong. After the war, there have been clear attempts by the government to revive falling agricultural production levels. The last 17 years have marked a gradual restructuring of agriculture given the introduction of new government agricultural policies.

In a new framework, the government of Kosovo has begun to formulate agricultural policies with a similarity in structure to those of the European Union. Despite the government's policy support, there are challenges that farmers face. There is still a strong dependence on imports of agricultural products and processed food (Sauer, Davidova, & Latruffe, 2012) which affects farmers' production levels and where farms in fact have been facing low production efficiency (Zuzaku, 2014). The production of greenhouse vegetables is a relevant example of a subsector that is experiencing low input efficiency use given its relatively new existence, alternative style of production, and the competition from imports.

There was little doubt that the war left agriculture in a seriously weakened economic state (Andersson, Rexhepi, Farinelli, & D'Costa, 2001). However, the resilience and the recent presence of greenhouse tomato and pepper farms marks a newly emerging form of production. The greenhouse farms in Kosovo are viewed comparatively advantageous relative to open field farms due to early season production. Yet, farmers of this fragile subsector argue that they are facing difficulties to efficiently incorporate proportionate input quantities in the production process.

In Kosovo, the Ministry of Agriculture, Forestry and Rural Development (MAFRD) has instituted funding initiatives. MAFRD (2016) introduced an investment activity titled "Measure 101: Investments in physical assets and agricultural households" which funded 65 out of 479 applications of farmers in 2015 with a value of 3,275,340.85 euros who applied to receive new greenhouses, support for open-field production and/or storage warehouses as farm facilities necessary in the production of vegetables (MAFRD, 2016). The structure of the financial support was primarily in the form of grants to promote the vegetable production. Without an empirical assessment, however, any conclusions reached about the effectiveness of MAFRD's financial support in the increase of input efficiency use can be misleading.

There is acknowledgment by MAFRD that growing seasons with low yields hamper farmers' ability to reduce input costs. The underlying tension between the associated aspects of low yields and the inefficient use of inputs may lead to a decrease in the domestic production of greenhouse tomatoes and peppers. Several studies have found the optimal use of inputs to be an important determinant of the vegetable production (Alboghdady, 2014; Nikolla, et al., 2013). While Kaciu, Babaj, Aliu, and Demaj (2016) have examined the drivers of vegetable production efficiency in Kosovo, there was little attention given to the greenhouse tomato and pepper input efficiency use. Moreover, maximizing input efficiency is a subject that has not been examined in detail. For this suite of reasons, to understand how to maintain current yields while decreasing waste of resources, input efficiency use can be studied.

**Purpose**

In developing countries like Kosovo, low production efficiency harms farmers' interests and may lead to a narrow path to achieve good farm decision making. The greenhouse vegetable sector was chosen because it represents the least scientifically explored and yet with the most economic potential for agriculture in Kosovo. A research-based analysis of input efficiency use may help farmers eliminate unnecessary costs and provide greenhouse tomatoes and peppers in the market with lower prices. This may be possible due to the identification of the optimal input values that may be used in the production process.

The major farm costs concerning farmers are the increasing costs of labor, pesticide, artificial fertilizer in the planting phase, and crystalline and artificial fertilizer in the flowering phase. Access to and utilization of these inputs with minimum costs in the production process pose a research dilemma, which to date has been little explored. The attention of this study is to examine the reduction of greenhouse inputs while maintaining tomato and pepper yields. A non-parametric application of input-oriented data envelopment analysis (DEA) is performed to identify the optimal values of inputs. To evaluate efficiency, an emphasis is put on the use of DEA's Charnes-Cooper-Rhodes (CCR) and Banker-Charnes-Cooper (BCC) models. A large body of literature has examined the application of DEA as an efficiency measurement method in agriculture (Adhikaria & Bjorndalb, 2012; Raheli, Rezaei, Jadidi, & Mobtaker, 2017). However, few researchers have examined its use for the greenhouse tomato and pepper input efficiency use, particularly in the context of Kosovo.

With the increasing evidence on the use of linear and logistic regression in agriculture (Battilani, et al., 2008; Huat, Doré, & Aubry, 2013; Chivuraise, Chamboko, & Chagwiza, 2016), this study includes an additional procedure to DEA which uses both models to determine the impact of other external, however, production related variables on the technical efficiency (TE) of the greenhouse tomatoes and peppers.

**Methods**

The purpose of this study was to determine the technical efficiency (TE), pure technical efficiency (PTE), and scale efficiency (SE) of the greenhouse tomato and pepper farms. An additional procedure was to analyze factors that influence TE scores. The input-oriented data envelopment analysis (DEA) coupled with multiple linear and binary logistic regression were utilized to compare input efficiency use between farms and to explore what farm variables could predict TE. To quantify input efficiency use, the input-oriented Banker-Charnes-Cooper (BCC) and Charnes-Cooper-Rhodes (CCR) models were used. The input-oriented direction of both models provided the possibility to investigate if input quantities may be reduced while keeping greenhouse tomato and pepper yields constant.

An emphasis was put on the compatible application of the linear and binary logistic regression together with input-oriented BCC and CCR models for detecting additional variable effects on the CCR efficiency. This was a second step to input-oriented DEA. For the second step analysis, the TE scores received from the greenhouse tomato and pepper CCR models were related to the variables which were considered to have an external impact on input efficiency use. It followed that model calculations were performed in R, a programming language and software for statistical analysis.

*Pure Technical Efficiency*

The BCC model originally introduced by Banker, Cooper, and Charnes (1984) and discussed more recently by Banker, Cooper, Seiford, Thrall, and Zhu (2004) was used for the evaluation of the greenhouse tomato and pepper farms' PTE. A tomato or pepper greenhouse was represented by n, and continued j = 1, 2, ..., n, obtained the same s yields, $y_{rj}$ (r = 1, 2, ..., s), using the same m farm inputs, $x_{ij}$ (i = 1, 2, ..., m). However, yields and farm inputs were in different amounts in this study given that each farmer reported a specific value of yields per growing season using varying amounts of inputs. In addition, epsilon (ε) denotes a small value to avoid categorizing an inefficient decision making unit (DMU) as efficient, and Korhonen & Joro (2015) argued that a small and positive value for ε may work in many instances. The BCC model as shown by Banker et al. (2004) was evaluated as follows.

$$\min \theta_o - \varepsilon \left( \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right),$$

subject to

$$\theta_{o}x_{io} = \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- \quad i = 1, 2, \ldots, m,$$

$$y_{ro} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ \quad r = 1, 2, \ldots, s,$$

$$1 = \sum_{j=1}^{n} \lambda_j,$$

$$0 \leq \lambda_j, s_i^-, s_r^+ \quad \forall\ i, r, j,$$

(1)

Yield as a main indicator of success may be considered as the output for both greenhouse tomato and pepper farms. Because farmers' main production decisions can be based on the potential future yields (Tomek & Kaiser, 2014). The inputs hypothesized to produce the best models were insecticide, labor, greenhouse area in square meters ($m^2$), greenhouse value in euros, as well as the use of artificial and organic fertilizers at different stages of greenhouse production. In Kosovo, often greenhouses covering larger areas correspond to greenhouses that have improved designs and structures. These greenhouses may have on average higher volumes of production and may be more efficient; therefore, the variable greenhouse area in $m^2$ was included.

Among pesticide use in Kosovar greenhouse tomato and pepper production, insecticides are often used more regularly than herbicides and fungicides. A further consideration for the inclusion of the insecticides was the study of Kaciu (2008) in the context of Kosovo, who stated that chemical measures are costly, however, with a high presence and potential to eliminate the insects. And insecticides may be highly beneficial to production. Labor was included as an input because two studies, Alboghdady (2014) and Zalkuw, Singh, Pardhi, and Gangwar (2014) presented evidence that labor is often overused in tomato production. Because labor has been shown as an input prone to overuse in greenhouse tomato production, it was hypothesized that it was also overused in pepper production.

The literature review noted that fertilizer was an important variable to be included in the evaluation of TE (Thimmareddy, Desai, & Vinoda Kumar, 2013; Zalkuw, et al., 2014). One consideration is whether artificial or organic fertilizer may impact more the greenhouse tomato and pepper yields. Although several studies noted that organic fertilizers may provide more increased productivity (see e.g. Mader et al., 2002; Dumas, Dadomo, Di Lucca, & Grolier, 2003; Sohail, 2008), still Heeb, Lundegårdh, Savage, and Ericsson (2006) determined that yield of tomatoes was higher with the use of mineral fertilizers compared to organic fertilizers. There are two reasons why it may be important to consider planting and flowering phase fertilizers in the evaluation of input efficiency use. First, they were expected to have an impact on yields, and their optimal use may ensure an increase in the production of greenhouse tomatoes and peppers. Second, an over or under utilization of these inputs may lead farms to incur production losses. To explore how fertilizers impact both greenhouse tomato and pepper input efficiency use, this study included the planting phase organic and artificial fertilizers, and flowering phase crystalline and artificial fertilizers.

In discussions with Kosovar experts, the value of greenhouses in Kosovo varied greatly between government program grantees and non-grantees. Grantees had greenhouses of higher euro value. Therefore, this study considered also the greenhouse value in euros as an input for the evaluation of input efficiency use among greenhouse tomato and pepper farms.

*Technical Efficiency*

The CCR model originally introduced by Charnes, Cooper, and Rhodes (1978) was as a continuation to the BCC model, and Banker et al. (2004) showed that it is in the same "envelopment form" of the BCC model except for the only omitted condition $\sum_{j=1}^{n} \lambda_j = 1$. This omitted condition noted that the BCC model may account for the variable returns to scale (VRS) assumption, while the CCR model with the condition of $0 \leq \lambda_j$ allowed for the constant returns to scale (CRS) assumption. This model using the same inputs and outputs explained previously served in this study for the evaluation of the greenhouse tomato and pepper farms' TE.

$$\min \theta - \varepsilon \left( \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right),$$

subject to

$$\theta_{X_{io}} = \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^-,$$

$$y_{ro} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+,$$

$$0 \leq \lambda_j, s_i^-, s_r^+ \quad \forall\ i, j, r. \tag{2}$$

*Scale Efficiency*

In a linear programming framework, DEA as a non-parametric method was used to calculate scale efficiency (SE). The evaluation of the CCR and BCC models helped us define SE as follows.

$$SE = {}^{\theta_{CCR}}/_{\theta_{BCC}} \tag{3}$$

The $\theta_{CCR}$ and $\theta_{BCC}$ represent the efficiency scores of a DMU or a greenhouse farm with the help of CCR under constant returns to scale (CRS) and BCC under variable returns to scale (VRS), respectively. For a BCC-efficient DMU with CRS characteristics, its SE is equal to one. The CCR efficiency score takes no account of scale effect under CRS and represents TE, while the BCC exhibits PTE under VRS. With the use of these notions, this study demonstrated the decomposition of efficiency:

$$\text{Technical Efficiency (TE)} = \text{Pure Technical Efficiency (PTE)} \times \text{Scale Efficiency (SE)} \tag{4}$$

The purpose of the decomposition was to show whether the source of inefficiency was caused by PTE, by SE, or by both (Cooper, Seiford, & Tone, 2002).

*Logistic Regression*

To develop a two-step analysis, this study applied first a multiple binary logistic regression as a multivariate statistical technique. The efficiency scores from the CCR model were selected as the dichotomous outcome. An outcome with a value equal to one suggested an efficient tomato or pepper greenhouse and zero otherwise. To investigate the research question which farm inputs reasonably predict greenhouse tomato CCR technical efficiency, this study used the multiple binary logistic regression.

Tomato Logistic Regression Model

$$P_j = {e^{Z_i}}/{e^{Z_i+1}} = {1}/{1+e^{-Z_i}} = E\left(Y_i \middle| \begin{array}{c} TCROP_i + TPOWER_i + TROWS_i + TWHOLE_i + TEXREV_i + TOTHER_i \\ + TFARM_i + TWELL_i + TWATER_i + TEDU_i + TFAMILY_i \end{array}\right)$$

Tomato Technical Efficiency as the Binary Outcome

$$\begin{aligned} Z_i &= \beta_0 + \beta_1 TCROP_i + \beta_2 TPOWER_i + \beta_3 TROWS_i + \beta_4 TWHOLE_i + \beta_5 TEXREV_i + \beta_6 TOTHER_i \\ &+ \beta_7 TFARM_i + \beta_8 TWELL_i + \beta_9 TWATER_i + \beta_{10} TEDU_i + \beta_{11} TFAMILY_i \end{aligned} \tag{5}$$

Under these circumstances, an evaluation including variables used for the greenhouse tomato farms from equation (5) may also affect the greenhouse pepper farms' optimal use of inputs. Production specifities of the greenhouse tomatoes and peppers in Kosovo may dictate the use of similar variables in the regression analyses. It may hold that in the context of vegetable production in Kosovo variables used in equation (5) may have an equal impact on the greenhouse pepper farms.

Pepper Logistic Regression Model

$$P_j = e^{Z_i} / e^{Z_i+1} = 1 / 1+e^{-Z_i} = E\left(Y_i \middle| \begin{array}{c} PCROP_i + PPOWER_i + PROWS_i + PWHOLE_i + PEXREV_i + POTHER_i \\ + PFARM_i + PWELL_i + PWATER_i + PEDU_i + PFAMILY_i \end{array}\right)$$

Pepper Technical Efficiency as the Binary Outcome

$$Z_i = \beta_0 + \beta_1 PCROP_i + \beta_2 PPOWER_i + \beta_3 PROWS_i + \beta_4 PWHOLE_i + \beta_5 PEXREV_i + \beta_6 POTHER_i$$
$$+ \beta_7 PFARM_i + \beta_8 PWELL_i + \beta_9 PWATER_i + \beta_{10} PEDU_i + \beta_{11} PFAMILY_i \quad (6)$$

In both models, $Y_i$ indicated if the tomato or pepper greenhouse was efficient or inefficient in the use of inputs. The outcome of the binary logistic regression was given by $P_j$ which modeled the probability of zero and one. Following this approach, this study used TE scores for $Z_i$ as the dependent variable for a given greenhouse tomato or pepper farm. A DMU with a CCR technical efficiency value of one also received a value of one in the logistic regression analysis and zero indicated otherwise. Coding a binary dependent variable based on the received efficiency score value of 1 and as zero otherwise was discussed by Tanfani and Testi (2012). For direct interpretation of the results, odds ratios of the coefficients were provided.

The variable T/PCROP was an indicator variable coded as one if a farmer expressed his need for a crop nutrition training and zero otherwise. In the literature, non-formal knowledge was found to attain higher technical efficiency and improve farm performance (see e.g. Manevska-Tasevska, 2013). It was useful to explore which group of farmers were more efficient in the use of inputs, those who expressed their need to participate in a crop nutrition training or those who did not.

Another indicator variable was T/PPOWER which had electricity coded as one and fuel otherwise. The government of Kosovo has prioritized the energy sector with an emphasis to improve electricity generation capacities. Over the years, a steady increase in electricity production has been noted in Kosovo, however, challenges remain for an efficient use of electricity (MED, 2017). Enterprises and farms in Kosovo may examine a cost-effective way to consume energy. There may be variations between greenhouse farmers who used electricity and greenhouse farmers who used fuel as the power source at a farm level. It may be vital to consider this variable's effect on greenhouse farms' TE. Still, electricity remains a costly source of power in Kosovo (Bowen, Myers, Myderrizi, Hasaj, & Halili, 2013). The prominence of the variable was related also to the fact that farmers focus on minimizing costs where the source of power often accounts for a large share in the vegetable production costs. More explanatory strength may be added to the model given that the optimal use of inputs may be affected by the source of power consumed.

The variable T/PROWS represented the total number of greenhouse tomato or pepper rows. Fruit yields may increase when in the greenhouse plants are arranged correctly and when there is a minimization of gaps between plants and rows (Rodriguez, Shaw, & Cantliffe, 2007). To achieve more greenhouse tomato or pepper yields, farmers may increase the number of rows more than it may be efficient. Likewise, a discrepancy in the number of rows per greenhouse may impact how each farmer uses inputs in the production process. It is of interest to find an appropriate number of rows in the greenhouse which may impact how inputs are allocated. The ability to research how this discrepancy affects the optimal use of inputs may justify the inclusion of the variable in the models.

An issue often reported from the greenhouse farmers is the low price received per kg of the produce. In this study, T/PWHOLE showed the wholesale price per kg of tomatoes or peppers received from the vegetable wholesalers. There is in fact a high volatility of prices among tomatoes (Alboiu, 2011). Farmers can have high price expectations, if they noticed that there were high wholesale prices in the market from the previous harvesting season

(Haile, Kalkuhl, & Usman, 2015). When the price is low from the vegetable wholesalers, greenhouse farmers may have to contract their gross profit margins. Farmers to avoid the risk of not selling may be even forced to market at lower prices their produce. Particularly, the wholesale price for tomatoes tends to fluctuate (Jaleta & Gardebroek, 2007). To test if the variable including farmer's wholesale selling price has an impact in the optimal use of inputs, this study included it in the models.

The variable T/PWATER indicated the irrigation equipment value in euros in the production of greenhouse tomatoes and peppers. The use of this variable was considered as a relevant way of understanding a farmer's quality of the irrigation system. Inadequate irrigation of the vegetable crops because of the old irrigation equipment may constrain the input efficiency use. Despite the wide presence of the drip irrigation systems in Kosovo, Balliu and Kaçiu (2008) stated that the frequency and amount of irrigation needs improvement. The greenhouse tomato and pepper farmers may have irrigation systems with varying euro values. Therefore, it was essential to consider this variable together for the greenhouse tomato and pepper input efficiency use related regression analyses.

This study included T/PEDU to indicate education in years for both the greenhouse tomato and pepper farmers. There are studies in agriculture that have found education to positively impact higher levels of TE (Balcombe, Fraser, Rahman, & Smith, 2006; Theodoridis & Anwar, 2011). However, Coelli, Rahman, and Thirtle (2002) noted that education may not be significantly correlated with efficiency. In this frame of analysis, this study included T/PEDU to test whether education has an impact on the greenhouse tomato and pepper TE.

The use of T/PFAMILY denoted a greenhouse tomato or pepper farmer's number of family members. Kaciu (2008) stated that most of the farm work is performed by the farmer's family members. In fact, the number of family members may dictate the intensity of the family labor use. The limited literature in Kosovo on the impact of the family members in the production of the greenhouse tomatoes and peppers allowed this study with its empirically-based analysis to examine any potential influence of the variable.

The variable T/PEXREV was an indicator variable coded as one if a farmer reported an external source of revenue and zero otherwise. Prices and varying yields often influence farm incomes (Barry, Hopkin, & Baker, 1988). In fact, off-farm income was found to have a positive effect on revenue risk (El Benni, Finger, & Mann, 2012). The use of T/PEXREV tested whether farmers with an external source of revenue were comparatively different to farmers without an external source of revenue in the optimal use of inputs. Off-farm income may even substitute for income losses that occurred in the farm (Blank & Erickson, 2007; El Benni, Finger, & Mann, 2012). However, there may not be sufficient evidence to conclude why some farmers rely on off-farm income and others do not (Blank & Erickson, 2007). It can be expected in this study that farmers who have an external source of revenue may rely less on on-farm revenue.

Another variable included in the model was T/POTHER, which indicated whether a farmer grew other crops in the greenhouse. This variable was coded as one if the farmer reported to have grown only greenhouse tomatoes or peppers. Depending on the number of other crops grown with tomatoes or peppers, this variable continued to take a maximum value of up to four. Vegetable farms growing two or more crops were found to have less usage of water, diesel and electricity (Li, et al., 2018). In Kosovo, there may be a mixture of farmers growing greenhouse tomatoes or peppers as a single crop and those who may have other crops with tomatoes or peppers in the same greenhouse. Błażejczyk-Majka, Kala, and Maciejewski (2012) stated that large-sized and mixed farms tend to have high efficiency. The inclusion of the T/POTHER may be to understand how growing other crops over the course of a season impacts the efficient use of inputs.

The variable T/PFARM represented the price received per kg of greenhouse tomatoes or peppers from the farmer's market. Given that product prices in agriculture have a high tendency to vary (Tomek & Kaiser, 2014), farmers may choose to sell directly to the retailers or consumers through farmers' markets (Ahearn & Sterns, 2013). In fact, high quality peppers can achieve premium prices in the market (Sephton, 2010). While, conventional tomatoes relative to organic tomatoes may not fetch premium prices in the market (Huang & Lin, 2007). There may be a belief in Kosovo that farmers selling greenhouse peppers at the farmer's market may receive higher prices per kg compared to the greenhouse tomato farmers. Farmers often sell at the farmer's market given the possibility to reach costumers directly. Whether the farmer market price influences the efficient use of inputs was of interest to explore.

Lastly, T/PWELL indicated well depth in meters. The amount of water applied on crops has a clear tendency to affect yields (Provenzano, Cots, Monserrat, Autovino, & Barragán, 2016), and an efficient use of irrigation would rely on the design of the irrigation system and its management (Barragan, Cots, Monserrat, Lopez, & Wu, 2010). For example, an implication of a limited irrigation time may suggest that farms in Kosovo's regions with lower average well depths may be less likely to irrigate during the flowering season when faced with increased levels of water scarcity. When the well depth is large and there is an increase in irrigation effectiveness, a potential to grow yields is possible (Caswell & Zilberman, 1986). This study expects that it may be possible to test if well depth is likely to affect the optimal use of inputs.

*Linear Regression*

To examine the average effect of the TE scores, a multiple linear regression was applied separately for the greenhouse tomato and pepper farms. For the scalar dependent variable, TE scores derived from the CCR model were used to understand the statistical impact of external production related variables. The process of using TE scores for both greenhouse tomato and pepper farms without a threshold may provide comparatively different results to the logistic regression analyses.

Tomato Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \beta_8 x_{8i} + \beta_9 x_{9i} + \beta_{10} x_{10i} + \beta_{11} x_{11i} + \varepsilon_i$$

Technical Efficiency as the Scalar Dependent Variable

$$TE_i = \beta_0 + \beta_1 TCROP_i + \beta_2 TPOWER_i + \beta_3 TROWS_i + \beta_4 TWHOLE_i + \beta_5 TEXREV_i + \beta_6 TOTHER_i + \beta_7 TFARM_i + \beta_8 TWELL_i + \beta_9 TWATER_i + \beta_{10} TEDU_i + \beta_{11} TFAMILY_i + \varepsilon_i \quad (7)$$

An evaluation of the same variables with the same justifications and definitions from equation (5) were performed for the linear regression model including greenhouse pepper farms. An additional comparison of the logistic regression to the linear regression in the prediction of input efficiency use was part of the study for both greenhouse tomato and pepper farms. While, the linear regression model for the greenhouse pepper farms was considered as follows.

Pepper Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \beta_8 x_{8i} + \beta_9 x_{9i} + \beta_{10} x_{10i} + \beta_{11} x_{11i} + \varepsilon_i$$

Technical Efficiency as the Scalar Dependent Variable

$$TE_i = \beta_0 + \beta_1 PCROP_i + \beta_2 PPOWER_i + \beta_3 PROWS_i + \beta_4 PWHOLE_i + \beta_5 PEXREV_i + \beta_6 POTHER_i + \beta_7 PFARM_i + \beta_8 PWELL_i + \beta_9 PWATER_i + \beta_{10} PEDU_i + \beta_{11} PFAMILY_i + \varepsilon_i \quad (8)$$

Following the models, this study called $TE_i$ the technical efficiency as the dependent variable for a given tomato or pepper greenhouse. The random error term for a greenhouse was represented by $\varepsilon_i$. The evaluation procedure included the same variables with the same justifications and definitions from equation (6) for the linear regression model including greenhouse pepper farms. For the linear regression models, a scalar dependent variable was used, and no threshold was performed to determine if a greenhouse was efficient or inefficient in the use of inputs based on the value it received in the CCR model.

## Data

### *Summary of Data*

The study's data were collected from June 1, 2017 to August 7, 2017 with a sample covering 136 greenhouse tomato and pepper farms from 22 villages, 11 municipalities, and 7 regions of Kosovo[1]. A special emphasis was put on regions that are characterized with greenhouse tomato and pepper production. Production information was obtained using face-to-face interviews. To facilitate the data collection process, two research surveys[2] were developed to gather information from the field in the regions of Prishtina, Ferizaj, Gjilan, Prizren, Gjakova, Peja, and Mitrovica. The first research survey was developed for the greenhouse tomato farms, and the second for the greenhouse pepper farms. Each research survey had 47 questions and separated into four survey modules. The first, second, and fourth modules named "The main respondent", "Greenhouse data", and "Greenhouse producer needs" had the same nature of questions across both surveys, while the third module "Greenhouse tomato/pepper farm data" was designed specifically for the tomato or pepper production, respectively.

Farmers interviewed were growing at least tomatoes or peppers as their cash crop. Some of those that cultivated tomatoes or peppers produced a few rows of these or other vegetables for home consumption. Although there were limited questions that asked for other

[1] This study during the field visits in the region of Prizren did not encounter commercial greenhouse pepper farms. Thus, Prizren is a missing region in the evaluation process.
[2] These surveys received approval from the University of Arkansas Institutional Review Board (Approval number 17-04-678) and were carried out with collaboration by faculty at the University of Prishtina "Hasan Prishtina" in Kosovo.

crops grown, it was outside the scope of this study to systematically collect data for non-commercial vegetables. The research survey was only applicable to farmers who grew vegetable crops for the market. Of all the farmers interviewed, 94 produced tomatoes and 42 produced peppers. Out of the 94 farmers producing tomatoes, there were seven farmers who produced both tomatoes and peppers.

*Variable Definitions and Descriptive Statistics*

Table 2.1 shows and defines eight inputs and one output described in the methods above which may be considered for evaluating greenhouse tomato input efficiency use and seven inputs and one output for the evaluation of pepper input efficiency use. In this farm survey sample, planting phase artificial fertilizer applied in kilograms (kg) was absent among greenhouse pepper farms. However, they did report the application of artificial fertilizer in the flowering phase.

Table 2.1. Description of Inputs and Outputs for Data Envelopment Analysis

| Variables | Description |
|---|---|
| *Tomato Inputs (x)* | |
| TOFERT | Planting phase organic fertilizer applied in kilograms |
| TAFERT | Planting phase artificial fertilizer applied in kilograms |
| TCFERT | Flowering phase crystalline fertilizer applied in kilograms |
| TEAFERT | Flowering phase artificial fertilizer applied in kilograms |
| TINSEC | Insecticides sprayed in liters |
| TLABOR | Combined family and hired labor as active working days per season |
| THOUSEVAL | Greenhouse value in euros |
| TAREA | Greenhouse area occupied with tomatoes in square meters |
| *Pepper Inputs (x)* | |
| POFERT | Planting phase organic fertilizer applied in kilograms |
| PCFERT | Flowering phase crystalline fertilizer applied in kilograms |
| PEAFERT | Flowering phase artificial fertilizer applied in kilograms |
| PINSEC | Insecticides sprayed in liters |
| PLABOR | Combined family and hired labor as active working days per season |
| PHOUSEVAL | Greenhouse value in euros |
| PAREA | Greenhouse area occupied with peppers in square meters |
| *Outputs (y)* | |
| TYIELD | Greenhouse tomato yields reported in kilograms |
| PYIELD | Greenhouse pepper yields reported in kilograms |

Table 2.2 gives the information concerning the inputs and outputs. Yield as a main indicator of success may be considered as the output for both greenhouse tomato and pepper farms. The average of the greenhouse tomato yields reported from farmers over the course of a growing season was 9,817 kilograms (kg), while the average of the greenhouse pepper yields was 4,648 kg. The input insecticide was the main chemical reported for the control of insects in the production process with an effect on the greenhouse tomato and pepper yields. A greenhouse tomato farmer sprayed an average of 149 liters of water with insecticides per growing season compared to a greenhouse pepper farmer with an average of 111 liters. However, there were farmers who applied zero amounts of water with insecticides, and farmers who applied as high as 800 liters among greenhouse tomato farmers and 400 liters among greenhouse pepper farmers.

The inputs showing fertilizer use suggested that on average greenhouse tomato farmers applied more planting phase artificial fertilizer and flowering phase crystalline fertilizer. However, greenhouse pepper farmers applied comparatively more planting phase organic fertilizer and flowering phase artificial fertilizer. The greenhouse tomato and pepper area in square meters ($m^2$) was an input with slight varying values. For example, a greenhouse tomato farmer had on average 50 $m^2$ more than greenhouse pepper farmers. The comparison is only for the area in $m^2$ occupied with tomatoes or peppers. The total greenhouse area in $m^2$ may exhibit higher variations. Examining the input consisting of the greenhouse value in euros suggested that on average greenhouse pepper farmers had greenhouses with a euro value of 9,930 higher than greenhouse tomato farmers. However, this may be misleading as there were some large greenhouse pepper farmers with indicatively high euro values reported for greenhouses. Therefore, the median may be a proper statistic to consider for this variable where greenhouse pepper farmers had greenhouses with a euro value of only 2,350 higher than greenhouse tomato farmers.

During the data collection process, there were mostly farmers who reported the use of family labor, and few farmers who had both family and hired labor. Larger greenhouse areas were reported to have higher levels of production and which often required a higher use of labor. On average, greenhouse tomato farmers reported to have a total of 92 active working days per growing season. An active working day consisted of a day when the farmer went to the greenhouse and worked full-time. While, greenhouse pepper farmers reported to have an average of 86 active working days per growing season.

Table 2.2. Descriptive Statistics of Inputs and Outputs

| Statistic | Unit | N | Mean | CV | Min | Median | Max |
|---|---|---|---|---|---|---|---|
| *Tomato inputs (x)* | | | | | | | |
| Insecticide | l | 94 | 149 | 1.28 | 0 | 90 | 800 |
| Labor | day | 94 | 92 | 0.30 | 25 | 88 | 153 |
| Greenhouse value | euro | 94 | 14,597 | 1.30 | 550 | 7,650 | 92,000 |
| Greenhouse area | $m^2$ | 94 | 693 | 1.04 | 100 | 500 | 4,200 |
| *Planting Phase Fertilizer:* | | | | | | | |
| Organic | kg | 94 | 6,799 | 1.07 | 0 | 5,600 | 40,000 |
| Artificial | kg | 94 | 43 | 2.93 | 0 | 0 | 450 |
| *Flowering Phase Fertilizer:* | | | | | | | |
| Crystalline | kg | 94 | 32 | 1.32 | 0 | 15 | 270 |
| Artificial | kg | 94 | 17 | 2.37 | 0 | 0 | 180 |
| *Pepper inputs (x)* | | | | | | | |
| Insecticide | l | 42 | 111 | 1.17 | 0 | 0 | 400 |
| Labor | day | 42 | 86 | 0.37 | 52 | 88 | 147 |
| Greenhouse value | euro | 42 | 16,729 | 1.43 | 1,500 | 10,000 | 150,000 |
| Greenhouse area | $m^2$ | 42 | 639 | 0.58 | 200 | 600 | 1,800 |
| *Planting Phase Fertilizer:* | | | | | | | |
| Organic | kg | 42 | 7,260 | 0.63 | 1,500 | 7,000 | 30,000 |
| *Flowering Phase Fertilizer:* | | | | | | | |
| Crystalline | kg | 42 | 22 | 1.11 | 0 | 19 | 90 |
| Artificial | kg | 42 | 42 | 1.76 | 0 | 0 | 300 |
| *Outputs (y)* | | | | | | | |
| Tomato yield | kg | 94 | 9,817 | 1.07 | 900 | 5,000 | 50,000 |
| Pepper yield | kg | 42 | 4,648 | 0.78 | 650 | 3,550 | 16,000 |

Note: N, number of observations; CV, coefficient of variation shown is defined as follows: $CV = \frac{\sigma}{\mu}$ which is presented as a coefficient in this study with σ indicating the covariate's standard deviation and μ the covariate's mean value. Note also that the variation in the greenhouse value is higher among greenhouse tomato farmers. Because only the outlier with a value of 150,000 affected the variation in the greenhouse value among greenhouse pepper farmers to become larger.

In the regression analyses, an equal number of external variables are considered for the greenhouse tomato and pepper input efficiency use. Table 2.3 shows and defines eleven independent variables and one dependent variable and one outcome for the greenhouse tomato and pepper farmers separately.

Table 2.3. Description of Regression Variables

| Variables | Description |
|---|---|
| *Tomato variables (X)* | |
| TCROP | 1 = for a crop nutrition training need, 0 = otherwise |
| TPOWER | 1 = electricity as the power source, 0 = for fuel |
| TROWS | Number of tomato rows per greenhouse |
| TWHOLE | Wholesale price per kilogram of tomatoes |
| TEXREV | 1 = for having external revenue, 0 = otherwise |
| TOTHER | 1 = for other greenhouse crops grown, 0 = otherwise |
| TFARM | Farmer market price per kilogram of tomatoes |
| TWELL | Well depth in meters |
| TWATER | Irrigation equipment value in euros |
| TEDU | Education in years |
| TFAMILY | Number of family members |
| *Pepper variables (X)* | |
| PCROP | 1 = for a crop nutrition training need, 0 = otherwise |
| PPOWER | 1 = electricity as the power source, 0 = for fuel |
| PROWS | Number of pepper rows per greenhouse |
| PWHOLE | Wholesale price per kilogram of peppers |
| PEXREV | 1 = for having external revenue, 0 = otherwise |
| POTHER | 1 = for other greenhouse crops grown, 0 = otherwise |
| PFARM | Farmer market price per kilogram of peppers |
| PWELL | Well depth in meters |
| PWATER | Irrigation equipment value in euros |
| PEDU | Education in years |
| PFAMILY | Number of family members |
| *Dependent variables (y)* | |
| TDEP | Tomato technical efficiency scores |
| PDEP | Pepper technical efficiency scores |
| *Binary outcomes (Y)* | |
| TOUT | Binary indicator of an efficient greenhouse |
| POUT | Binary indicator of an efficient greenhouse |

Table 2.4 shows a summary of the statistics for the possible choice of independent and dependent variables[3] for the linear regression and the independent variables and outcomes[4] for the binary logistic regression. The variables for the regression models may be different from data envelopment analysis (DEA). The objective was to estimate an additional effect of other production related variables present in the dataset on the optimal use of inputs. For the regression analyses, this study considered a selection of the variables that characterize greenhouse tomato and pepper farms within the vegetable production context in Kosovo.

When observing the variable education, greenhouse pepper farmers appeared to have on average more years of education than the greenhouse tomato farmers. There were greenhouse tomato and pepper farmers with as low as 8 years of education, and as high as 20 years of education. Examining the variable including the farmer's number of family members, it suggested that there was on average a slight difference in the number of family members between greenhouse tomato and pepper farmers. The former group of farmers had an average of 9 family members, while the latter an average of 8 family members, respectively. Another variable of interest was whether a farmer expressed his interest to participate in a crop nutrition training. Among greenhouse tomato farmers, 83 percent strongly agreed to participate in a crop nutrition training. Likewise, there were 86 percent of the greenhouse pepper farmers who strongly agreed to participate in a crop nutrition training. Percentage wise, both groups of farmers showed an interest to learn more about the nutritional needs of tomatoes and peppers.

The variable power source suggested that 58 percent of the greenhouse tomato farmers chose electricity over fuel relative to the 67 percent of the greenhouse pepper

---

[3] A scalar dependent variable was part of the linear models where the efficiency scores were received from DEA's CCR models and used without a threshold.

[4] The binary outcome was performed with a threshold which characterized an efficient or inefficient greenhouse farm and was part of the logistic regression models. If a farm was efficient in DEA's CCR model, it received an efficient value of one, and zero otherwise.

farmers. The percent discrepancy for the greenhouse tomato farms in the use of electricity or fuel may be an issue of significance to explore. An additional variable was the number of rows per greenhouse at the time of the interview. In fact, the number of rows per greenhouse was often influenced from the greenhouse type and region among greenhouse tomato and pepper farms. According to the greenhouse type, farmers reported a high or low number of rows grown with tomatoes and peppers. This study's sample of greenhouse tomato farmers noted that a farmer having a multi tunnel greenhouse had an average of 19 rows, while a farmer with a single tunnel greenhouse had an average of 9 rows. Farmers from the region of Prizren had an aggregate mean of 17 rows per greenhouse. While, Mitrovica and Gjilan had aggregate means of 6 and 7 rows per greenhouse, respectively. When considering greenhouse pepper farmers, this study observed that farmers having a multi tunnel greenhouse had an average of 14 rows, while there were only 8 rows for a farmer growing peppers in a single tunnel greenhouse. At a regional level, there were farmers with large-sized greenhouses in Gjakova who had as high as 23 rows per greenhouse, and small-sized farmers as low as 4 rows per greenhouse in Gjilan.

The variable wholesale price per kg of tomatoes and peppers was of research interest. Greenhouse tomato farmers from Gjakova in certain situations in the market received an aggregate mean price of 0.10 euros per kg of tomatoes from the vegetable wholesalers, while greenhouse pepper farmers in some isolated cases received an aggregate mean price of as low as 0.09 euros per kg of peppers. There was only a limited number of greenhouse pepper farmers in Peja with low quality and late-season production who may have received a minimum aggregate mean price of 0.09 euros. However, it is important to note that most of the greenhouse pepper farmers in the region of Peja and other regions reported that they sold their produce to the farmer's market. There may be a possibility that the price received per kg of tomatoes or peppers affects the farmer's gross margins.

There was a slight mixture of farmers having external revenue and farmers having only the farm work as a source of revenue among the greenhouse tomato and pepper farms. The indicator variable showed that 27 percent of the greenhouse tomato farmers reported to have an external source of revenue, while only 26 percent among greenhouse pepper farmers. The variable including the farmer's market price per kg of tomatoes and peppers revealed from the dataset that at least 45 percent of the greenhouse pepper farmers were selling peppers at the farmer's market. In this line of reasoning, there were only 40 percent of the greenhouse tomato farmers selling tomatoes to the farmer's market. A greenhouse pepper farmer received an average price of 0.55 euros per kg of peppers, while a greenhouse tomato farmer received a price of 0.14 euros per kg of tomatoes. In addition, the variable if other crops were grown in the greenhouse together with peppers or tomatoes showed the number of crops grown in the farm. From the 42 surveyed greenhouse pepper farmers, there were farmers who grew three other vegetable crops with peppers. And, there were 5 percent of the greenhouse pepper farmers who grew only peppers in the greenhouse. When examining the 94 surveyed greenhouse tomato farmers, there were 55 percent who grew only one other vegetable crop with greenhouse tomatoes, while only 20 percent grew tomatoes as a single crop.

When lastly examining well depth in meters and irrigation equipment value in euros, it was suggested that for both variables greenhouse tomato farmers had higher varying values relative to greenhouse pepper farmers. The location of the farm may dictate in part the depth of the well and water availability. Nevertheless, a lack of knowledge for crop irrigation requirements in certain cases may also influence the farmer's decision to have a certain depth of the well and the quantity of water for irrigation.

Table 2.4. Descriptive Statistics of Regression Variables

| Statistic | Unit | N | Mean | CV | Min | Median | Max |
|---|---|---|---|---|---|---|---|
| *Tomato variables (X)* | | | | | | | |
| Crop nutrition training | 0/1 | 94 | 0.83 | 0.46 | 0 | 1 | 1 |
| Power source | 0/1 | 94 | 0.59 | 0.85 | 0 | 1 | 1 |
| Rows per greenhouse | number | 94 | 12.78 | 0.76 | 4 | 10 | 56 |
| Wholesale price | euro | 94 | 0.18 | 0.92 | 0 | 0.20 | 0.45 |
| External Revenue | 0/1 | 94 | 0.27 | 1.66 | 0 | 0 | 1 |
| Farmer market price | euro | 94 | 0.14 | 1.63 | 0 | 0 | 0.70 |
| Other crops grown | number | 94 | 2.11 | 0.36 | 1 | 2 | 4 |
| Well depth | m | 94 | 9.01 | 0.38 | 4 | 8 | 18 |
| Irrigation equipment value | euro | 94 | 463 | 1.04 | 10 | 400 | 3,000 |
| Education | years | 94 | 11 | 0.30 | 8 | 8 | 20 |
| Family members | number | 94 | 9 | 0.48 | 4 | 8 | 33 |
| *Pepper variables (X)* | | | | | | | |
| Crop nutrition training | 0/1 | 42 | 0.86 | 0.41 | 0 | 1 | 1 |
| Power source | 0/1 | 42 | 0.67 | 0.72 | 0 | 1 | 1 |
| Rows per greenhouse | number | 42 | 10.79 | 0.72 | 4 | 8.5 | 45 |
| Wholesale price | euro | 42 | 0.09 | 2.57 | 0 | 0 | 0.80 |
| External revenue | 0/1 | 42 | 0.26 | 1.70 | 0 | 0 | 1 |
| Farmer market price | euro | 42 | 0.55 | 1.15 | 0 | 0 | 1.70 |
| Other crops grown | number | 42 | 2.52 | 0.33 | 1 | 3 | 4 |
| Well depth | m | 42 | 8.59 | 0.33 | 3 | 8.50 | 16 |
| Irrigation equipment value | euro | 42 | 491 | 0.59 | 5 | 500 | 1,500 |
| Education | years | 42 | 12 | 0.27 | 8 | 12 | 20 |
| Family members | number | 42 | 8 | 0.34 | 4 | 7 | 19 |
| *Dependent variables (y)* | | | | | | | |
| TDEP | number | 94 | 0.47 | 0.61 | 0.14 | 0.38 | 1 |
| PDEP | number | 42 | 0.67 | 0.41 | 0.19 | 0.68 | 1 |
| *Binary outcomes (Y)* | | | | | | | |
| TOUT | 0/1 | 94 | 0.16 | 2.31 | 0 | 0 | 1 |
| POUT | 0/1 | 42 | 0.26 | 1.70 | 0 | 0 | 1 |

Note: N, number of observations; CV, coefficient of variation shown is defined as follows: $CV = \frac{\sigma}{\mu}$ which is presented as a coefficient in this study with σ indicating the covariate's standard deviation and μ the covariate's mean value. Price-related variables with a minimum or median of zero indicate farmers who did not sell to a vegetable wholesaler or farmer's market. Their value of zero was known, therefore, they were not assigned a value showing missing data.

## Results

### *Overall Input Efficiency Use*

Results of the data envelopment analysis (DEA) highlighted differences in input efficiency use for the greenhouse tomato and pepper farms in Kosovo. The production specifities were found to be different among greenhouse tomato and pepper farms. Based on the statistical significance of the variables and the literature review, an optimal and slightly different variable set was found for the greenhouse tomato and pepper farms' input efficiency analyses.

The final estimation procedure for tomatoes and peppers varied slightly between crops. The estimation procedure for the greenhouse tomato farms included yield in kilograms (kg) as the output, and insecticide, labor, greenhouse area in square meters ($m^2$), greenhouse value in euros and planting phase organic and artificial fertilizers, and flowering phase crystalline and artificial fertilizers as the inputs. The estimation procedure for the greenhouse pepper farms included yield in kg as the output, and insecticide, labor, greenhouse area in $m^2$, and planting phase organic, flowering phase crystalline and artificial fertilizers as the inputs. That is, greenhouse value in euros and planting phase artificial fertilizers were used only for tomatoes while all other inputs were used in both procedures. Using these variables, the Banker-Charnes-Cooper (BCC) model suggested more efficient greenhouse pepper farms relative to the greenhouse tomato farms with an absolute difference of 19.4 percent in the optimal use of inputs. While, the Charnes-Cooper-Rhodes (CCR) model also indicated that the greenhouse pepper farms were more efficient in the use of inputs with an absolute difference of 10.1 percent from greenhouse tomato farms.

The first step analysis provided an important suggestion that the greenhouse farms may increase the production of both vegetable crops through more efficient use of inputs. Under the variable returns to scale (VRS) technology, greenhouse tomato farms on average may improve their use of inputs by 24 percent, while under the constant returns to scale (CRS) technology, they may improve their use of inputs by as much as 53 percent. For the greenhouse pepper farms, an average improvement of 10 percent was suggested under the VRS technology and 33 percent under the CRS technology. Therefore, it may be likely that both greenhouse tomato and pepper farmers may use inputs more efficiently during the production process. Comparatively, the greenhouse pepper farms were more efficient than the greenhouse tomato farms. The former had an overall mean efficiency of 0.90, while the latter a mean efficiency of 0.76 under the VRS technology, respectively. Likewise, under the CRS technology, greenhouse pepper farms maintained a 0.67 mean efficiency relative to a 0.47 mean efficiency of the greenhouse tomato farms.

Often greenhouse vegetable farm differences in production have been considered as an explanation for the low efficiency. This may be because of the low production of vegetables and its inability to meet the domestic demand (Kosovo Report, 2006). However, given Kosovo’s approximate uniformity in weather conditions and cultivation practices, greenhouse tomato and pepper farms may not have varying production specifities that justify the over or under utilization of inputs. Prior to the detailed discussion of the results, this study noted that efficiency scores under the VRS technology were more relevant compared to the CRS technology only when comparing input efficiency use from region to region. The explanation was that CRS technology is a pertinent measure when all decision making units (DMUs) are operating at an optimal scale (Coelli, 1996). However, it was improbable in Kosovo’s context for all the greenhouse tomato and pepper farms to operate at an optimal scale. The analysis under the VRS technology thereby was of interest in this study for the comparison of input efficiency use at a regional level only. At a farm level, however, both technologies played a vital role to address sources of inefficiency.

*Greenhouse Tomato Input Use at a Farm Level*

Table 2.5 shows a summary of the range in greenhouse tomato efficiency, number of farms, and the percentage of farms for each efficiency range under the VRS and CRS technologies. The input-oriented DEA analysis of 94 greenhouse tomato farms noted inefficiency was present in the use of inputs. According to the BCC model, greenhouse tomato farms' pure technical efficiency (PTE) showed that 33 percent of the farms had an efficiency of 1.00. Farms with an efficiency score of 1.00 were using optimally their inputs, and the results may recommend no further changes to their performance. However, the 33 percent of the efficient farms under the BBC model implied that the remaining farms showed a tendency to have lower input efficiency use. While the CCR model for greenhouse tomato farms' technical efficiency (TE) suggested that only 16.1 percent of the farms had an efficiency of 1.00. Under this model, most farms were concentrated on the lower efficiency levels. Cooper, Seiford, and Tone (2002) stated that a DMU at the same time can be BCC-efficient and may be found CCR-inefficient from the CCR model. Depending on the objective of the researcher, more emphasis may be put on the analysis including BCC-efficient or CCR-efficient DMUs.

Table 2.5. Summary of Greenhouse Tomato Farms' Efficiencies

| BCC: Variable Returns to Scale Technology | | | CCR: Constant Returns to Scale Technology | | |
|---|---|---|---|---|---|
| E Range | # of farms | % | E Range | # of farms | % |
| 0.2<= E <0.3 | 1 | 1.1 | 0.1<= E <0.2 | 7 | 7.4 |
| 0.3<= E <0.4 | 1 | 1.1 | 0.2<= E <0.3 | 29 | 30.9 |
| 0.4<= E <0.5 | 9 | 9.6 | 0.3<= E <0.4 | 14 | 14.9 |
| 0.5<= E <0.6 | 15 | 16.0 | 0.4<= E <0.5 | 12 | 12.8 |
| 0.6<= E <0.7 | 15 | 16.0 | 0.5<= E <0.6 | 7 | 7.4 |
| 0.7<= E <0.8 | 11 | 11.7 | 0.6<= E <0.7 | 5 | 5.3 |
| 0.8<= E <0.9 | 7 | 7.4 | 0.7<= E <0.8 | 2 | 2.1 |
| 0.9<= E <1 | 4 | 4.3 | 0.8<= E <0.9 | 1 | 1.1 |
| E = 1 | 31 | 33.0 | 0.9<= E <1 | 2 | 2.1 |
| | | | E = 1 | 15 | 16.1 |
| Total | 94 | 100.2 | | 94 | 100.1 |

Note: E, efficiency; #, number; %, percentage. The technology of VRS with input-oriented efficiency had a mean of 0.76, minimum of 0.24, first quartile of 0.57, median of 0.75, third quartile of 1.00, and maximum of 1.00. The technology of CRS with input-oriented efficiency had a mean of 0.48, minimum of 0.14, first quartile 0.23, median of 0.38, third quartile of 0.63, and maximum of 1.00.

The comparison of the BCC and CCR models reveals the source of inefficiency. Of all the greenhouse tomato farms, 16 percent were scale efficient; this indicates that any changes to their existing input values may not lead to higher efficiency. It was suggested for scale efficient farms to maintain the level of efficiency through employing the same practices in the production process. These practices may further support the continuous and optimal use of inputs. However, 84 percent were facing decreasing returns to scale (DRS). Among the DRS farms, the results indicated that there may be an over utilization of the planting phase organic and artificial fertilizers, and flowering phase crystalline and artificial fertilizers compared to the scale efficient farms. Table 2.6 while presenting the mean and the coefficient of variation of each input under SE and DRS, it indicated that greenhouse tomato farms operating under DRS were oversized. In contrast, in terms of yields, scale efficient farms were slightly more than twice as efficient as their DRS counterparts.

Table 2.6. Greenhouse Tomato Efficient Input Values

| | | SE (N = 15) | | DRS (N = 79) | |
|---|---|---|---|---|---|
| Materials | Unit | Mean | CV | Mean | CV |
| *Inputs* | | | | | |
| Insecticide | l | 107 | 1.51 | 157 | 1.25 |
| Labor | days | 70 | 0.40 | 96 | 0.27 |
| Greenhouse area | $m^2$ | 955 | 1.16 | 644 | 0.96 |
| Greenhouse value | euro | 16,733 | 1.66 | 14,191 | 1.20 |
| *Planting phase fertilizer:* | | | | | |
| Organic | kg | 5,433 | 1.24 | 7,058 | 1.05 |
| Artificial | kg | 23 | 1.61 | 47 | 2.91 |
| *Flowering phase fertilizer:* | | | | | |
| Crystalline | kg | 12 | 6.5 | 36 | 1.22 |
| Artificial | kg | 0 | 0 | 21 | 2.10 |
| *Output* | | | | | |
| Yield | kg | 20,673 | 0.78 | 7,756 | 0.98 |

Note: SE, scale efficiency; DRS, decreasing returns to scale. Increasing returns to scale (IRS) were absent as greenhouse tomato farms experienced only SE or DRS. Coefficient of variation (CV) shown is defined as follows: $CV = \frac{\sigma}{\mu}$ which is presented as a coefficient in this study with $\sigma$ indicating the covariate's standard deviation and $\mu$ the covariate's mean value. Note also that yield is the total amount of tomatoes received in kg over the course of a growing season. There were large-sized farms with tomato yields of almost 30 kg per square meter ($m^2$) over the whole growing season, and small-sized farms with yields as low as 4-7 kg per $m^2$.

According to the results, greenhouse tomato farms facing DRS or diseconomies of scale may choose to reduce the farm output as they have surpassed their optimal size and use of inputs. In fact, the quantities of used inputs may affect the levels of yields (FAO, 2017).

Farmers interviewed reasoned that given the lack of production expertise among other factors for the optimal use of inputs in the production of greenhouse tomatoes, they were struggling to maintain yields without an increase of the inputs consumed. The results of this study reflected complaints from farmers that additional use of inputs may not provide higher yields particularly among the DRS farms.

To further understand the efficiency positioning of the farms, Figure 2.1 shows greenhouse tomato farms' efficiency frontiers under VRS using the BCC model, and CRS using the CCR model. The estimation procedure for the efficiency frontiers included yield as the output, and insecticide, labor, greenhouse area in $m^2$, greenhouse value in euros and planting phase organic and artificial fertilizers, and flowering phase crystalline and artificial fertilizers as the inputs. Greenhouse tomato farms on the frontier were those that used inputs most efficiently in the production process. The less efficient farms that were not on the frontier received efficiency scores lower than 1.00.

From Figure 2.1, a point on the piecewise linear curve represented a farm with an efficiency score of 1.00 from the BCC model, and the few points on the upward sloping line indicated efficiency values of 1.00 with zero-slack received from the CCR model. An efficiency score of 1.00 demonstrated a BCC- or CCR-efficient greenhouse tomato farm, respectively. However, when a DMU (farm) was found CCR-efficient, it implied that it will also be found efficient with the BCC model (Cooper, Seiford, & Tone, 2002). In addition, the gap in yields between small-sized and large-sized greenhouse tomato farms indicated the varying efficiencies of the two groups depicted in the graph. The first group consisted of most of the small-sized farms with similar yields demonstrating in some instances similar efficiency scores. They were concentrated at the bottom of the figure and some of the farms were overlapped. On the other extreme, a presence of large-sized farms scattered to the right with different yields were exhibiting varying efficiency scores. In contrast to the small-sized farms, large-sized farms appeared to have a disparity in the input intensity use. Overall, the former group of farms were utilizing inputs with almost the same manner with few exceptions. However, the latter group of farms have stark differences in the use of inputs.

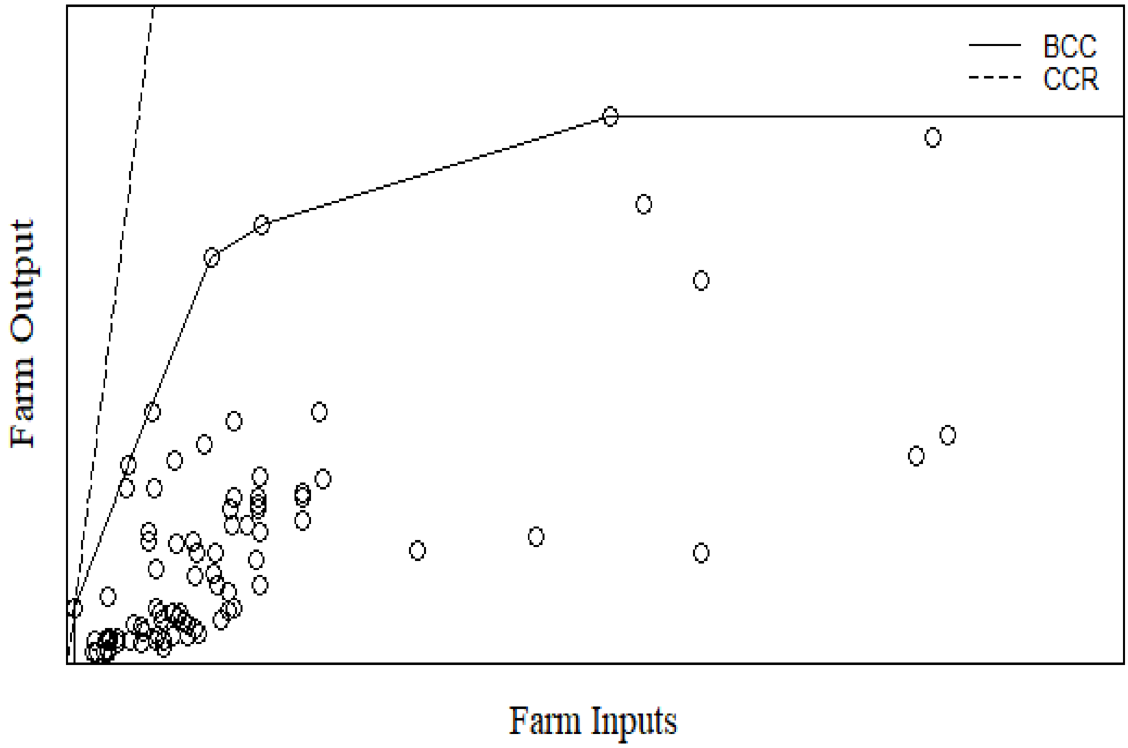


Figure 2.1. Greenhouse Tomato Farms' Efficiency Frontiers.
The BCC model represents the VRS technology with the piecewise linear curve, and the CCR model represents the CRS technology with the upward sloping line. Note that the plot includes multiple inputs. Therefore, aggregation occurred when attempting to plot all the BCC- or CCR-efficient farms on the frontiers using the Benchmarking package in R.

*Greenhouse Tomato Input Use Comparison at a Regional Level*

In addition to analyzing efficiency at a farm level, the extent of input efficiency use at a regional level was a component of the study. When using aggregate efficiency scores with VRS technology and BCC input-oriented efficiency, Prizren emerged as the region with the most efficient producers with a mean efficiency of 0.83. This result was in accordance with the study’s expectations. Because during the interview process, Prizren comprised the highest number of greenhouse tomato producers and with a preserved family farm tradition. The region of Prishtina followed with a mean efficiency of 0.80 for the optimal use of inputs in the production of greenhouse tomatoes. Under the CRS technology and CCR input-oriented efficiency, region of Peja with a mean efficiency of 0.58 ranked first, followed by Prizren with a mean efficiency of 0.56. Efficiency ranking of the regions under VRS was of relevance as CCR assumed all greenhouse tomato farms operate at an optimal scale. This assumption may not hold in Kosovo's newly emerging greenhouse sector. In this line of logic, the results indicating Prizren as the region with the most efficient producers were more likely to be consistent with the input efficiency use comparison at a regional level.

Table 2.7 provides an analysis including aggregate efficiency scores which were calculated as the mean efficiency in the use of inputs of a given region. The analysis including the aggregate efficiency scores allowed an observation of how the optimal use of inputs differs at a regional level. For a complete analysis, the minimum, maximum and the standard deviation were provided under PTE, TE, and SE for each region. By comparing the minimum and maximum of the PTE, TE, SE scores at a regional level, an understanding may be achieved as how far farms rank in input efficiency use for each region. A low minimum may indicate that there were producers that were not operating optimally. The use of standard deviation showed how the dispersion changes in input efficiency use among the greenhouse tomato producing regions. To complement this understanding, the exact number of farms falling under PTE, TE, SE may be found through observing the efficiency score range. The most regions under the analysis of PTE had significantly more efficient producers in the use of inputs compared to the analysis including TE. Earlier, it was indicated that PTE is of relevance in the comparison of efficiency between regions. The explanation was that TE originating from the CRS technology may be a pertinent measure when all DMUs are operating at an optimal scale (Coelli, 1996). Taken all greenhouse tomato farms together, it may be unlikely that each farm operates at an optimal scale. Figure 2.2 gives an efficiency representation of the leading greenhouse tomato producing regions under the VRS and CRS technologies. Farms with low efficiency may be observed in the lower quartile compared to the high efficiency farms in the upper quartile. However, this study underscored the mean efficiency of the regions, and how each region differs in the use of inputs according to this statistic.

Table 2.7. Greenhouse Tomato BCC, CCR, and Scale Efficiency Results

| Aggregate efficiency scores | | | | Efficiency score range and no. of farms | | | |
|---|---|---|---|---|---|---|---|
| Region | PTE | TE | SE | Range | PTE | TE | SE |
| *Prishtina* | | | | *Prishtina* | | | |
| Mean | 0.80 | 0.50 | 0.63 | <0.40 | 0 | 8 | 7 |
| Min | 0.51 | 0.17 | 0.26 | 0.40-0.69 | 7 | 8 | 5 |
| Max | 1.00 | 1.00 | 1.00 | 0.70-0.99 | 6 | 3 | 4 |
| St. Dev. | 0.17 | 0.26 | 0.27 | 1.00 | 6 | 0 | 3 |
| *Ferizaj* | | | | *Ferizaj* | | | |
| Mean | 0.59 | 0.36 | 0.63 | <0.40 | 0 | 11 | 1 |
| Min | 0.44 | 0.24 | 0.38 | 0.40-0.69 | 12 | 4 | 10 |
| Max | 0.77 | 0.52 | 0.89 | 0.70-0.99 | 3 | 0 | 4 |
| St. Dev. | 0.11 | 0.09 | 0.15 | 1.00 | 0 | 0 | 0 |
| *Gjilan* | | | | *Gjilan* | | | |
| Mean | 0.77 | 0.42 | 0.54 | <0.40 | 0 | 4 | 1 |
| Min | 0.43 | 0.20 | 0.29 | 0.40-0.69 | 4 | 4 | 5 |
| Max | 1.00 | 0.71 | 0.72 | 0.70-0.99 | 2 | 1 | 3 |
| St. Dev. | 0.24 | 0.20 | 0.16 | 1.00 | 3 | 0 | 0 |
| *Prizren* | | | | *Prizren* | | | |
| Mean | 0.83 | 0.56 | 0.65 | <0.40 | 2 | 17 | 15 |
| Min | 0.24 | 0.18 | 0.21 | 0.40-0.69 | 8 | 3 | 2 |
| Max | 1.00 | 1.00 | 1.00 | 0.70-0.99 | 6 | 2 | 5 |
| St. Dev. | 0.22 | 0.38 | 0.33 | 1.00 | 18 | 12 | 12 |
| *Gjakova* | | | | *Gjakova* | | | |
| Mean | 0.75 | 0.21 | 0.29 | <0.40 | 0 | 4 | 4 |
| Min | 0.60 | 0.14 | 0.16 | 0.40-0.69 | 2 | 0 | 0 |
| Max | 0.89 | 0.23 | 0.37 | 0.70-0.99 | 2 | 0 | 0 |
| St. Dev. | 0.14 | 0.05 | 0.10 | 1.00 | 0 | 0 | 0 |
| *Peja* | | | | *Peja* | | | |
| Mean | 0.70 | 0.58 | 0.84 | <0.40 | 0 | 1 | 0 |
| Min | 0.46 | 0.37 | 0.75 | 0.40-0.69 | 4 | 4 | 0 |
| Max | 1.00 | 0.83 | 0.95 | 0.70-0.99 | 1 | 2 | 7 |
| St. Dev. | 0.22 | 0.17 | 0.08 | 1.00 | 2 | 0 | 0 |
| *Mitrovica* | | | | *Mitrovica* | | | |
| Mean | 0.78 | 0.29 | 0.37 | <0.40 | 0 | 5 | 3 |
| Min | 0.52 | 0.19 | 0.30 | 0.40-0.69 | 2 | 1 | 3 |
| Max | 1.00 | 0.45 | 0.45 | 0.70-0.99 | 2 | 0 | 0 |
| St. Dev. | 0.21 | 0.10 | 0.06 | 1.00 | 2 | 0 | 0 |

Note: PTE, pure technical efficiency; TE, technical efficiency; SE, scale efficiency.

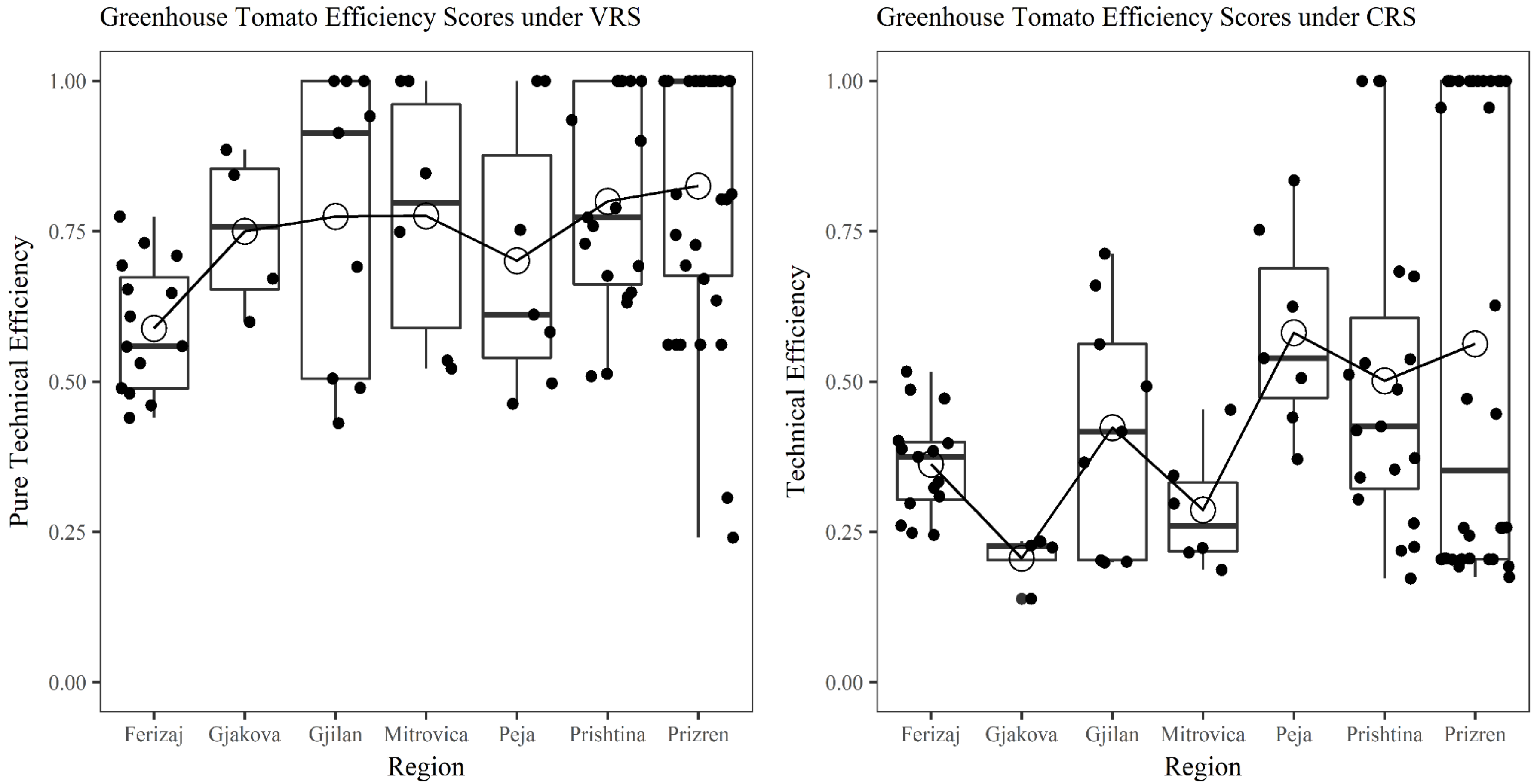


Figure 2.2. Efficiency Representation of Greenhouse Tomato Producing Regions
Note: Circles connected by lines show each region’s aggregate mean efficiency. Statistics presented are minimum, first quartile, mean, median, third quartile, and maximum. Under the CRS technology notice that region of Gjakova has a low standard deviation, and nearly the same median and maximum value which dictates the shape of the boxplot.

*Greenhouse Pepper Input Use at a Farm Level*

The results revealed that there were percentage wise less inefficient greenhouse pepper producers[5] compared to the greenhouse tomato producers at the lower efficiency levels. Table 2.8 shows a summary of the greenhouse pepper efficiency range, number of farms, and the percentage for each efficiency range under the VRS and CRS technologies. Greenhouse pepper farms' PTE showed that 52.4 percent of the farmers were BCC-efficient. While, the analysis including TE was lower with a 26.2 percent of the farmers as CCR-efficient.

Table 2.8. Summary of Greenhouse Pepper Farms' Efficiencies

| BCC: Variable Returns to Scale Technology | | | CCR: Constant Returns to Scale Technology | | |
|---|---|---|---|---|---|
| E Range | # of farms | % | E Range | # of farms | % |
| 0.3<= E <0.4 | 1 | 2.4 | 0.1<= E <0.2 | 1 | 2.4 |
| 0.4<= E <0.5 | 0 | 0.0 | 0.2<= E <0.3 | 4 | 9.5 |
| 0.5<= E <0.6 | 1 | 2.4 | 0.3<= E <0.4 | 4 | 9.5 |
| 0.6<= E <0.7 | 1 | 2.4 | 0.4<= E <0.5 | 6 | 14.3 |
| 0.7<= E <0.8 | 7 | 16.7 | 0.5<= E <0.6 | 1 | 2.4 |
| 0.8<= E <0.9 | 6 | 14.3 | 0.6<= E <0.7 | 6 | 14.3 |
| 0.9<= E <1 | 4 | 9.5 | 0.7<= E <0.8 | 5 | 11.9 |
| E = 1 | 22 | 52.4 | 0.8<= E <0.9 | 2 | 4.8 |
| | | | 0.9<= E <1 | 2 | 4.8 |
| | | | E = 1 | 11 | 26.2 |
| Total | 42 | 100.1 | | 42 | 100 |

Note: E, efficiency; #, number; %, percentage. The technology of VRS with input-oriented efficiency had a mean of 0.90, minimum of 0.40, first quartile of 0.81, median of 1.00, third quartile of 1.00, and maximum of 1.00. The technology of CRS with input-oriented efficiency had a mean of 0.67, minimum of 0.19, first quartile 0.45, median of 0.68, third quartile of 0.99, and maximum of 1.00.

Greenhouse pepper SE results showed that 26 percent of the farms were scale efficient, indicating a 10 percent higher SE compared to the greenhouse tomato farms. Under the scale of production, 7 percent were operating in the area of increasing returns to scale

[5] During the data collection process there were only few and often non-commercial greenhouse pepper farms in the region of Prizren. Therefore, greenhouse pepper farmers from the region of Prizren were absent in this study. However, open-field pepper production may be common and with a high number of pepper farmers.

(IRS), and 67 percent in the area of DRS. These results demonstrated that the production scale of the greenhouse pepper farms was small. Nevertheless, there was a slightly lower presence of diseconomies of scale compared to the greenhouse tomato farms. In this situation, farms were mostly overutilizing inputs without a corresponding increase in yields. For farmers operating under IRS, they may increase the use of inputs examined in this study to progress towards SE. In fact, scale efficient farms in terms of yields were slightly less than twice more efficient than their DRS and IRS counterparts. Table 2.9 presents efficient values of inputs under SE that DRS and IRS farms throughout Kosovo may consider for a comparison to their existing input quantities used in the production of greenhouse peppers.

Table 2.9. Greenhouse Pepper Efficient Input Values

| | | SE (N = 11) | | IRS (N = 3) | | DRS (N = 28) | |
|---|---|---|---|---|---|---|---|
| Materials | Unit | Mean | CV | Mean | CV | Mean | CV |
| *Inputs* | | | | | | | |
| Insecticide | l | 100 | 1.26 | 67 | 1.72 | 119 | 1.13 |
| Labor | days | 80 | 0.19 | 82 | 0.22 | 90 | 0.24 |
| Greenhouse area | $m^2$ | 745 | 0.57 | 383 | 0.08 | 624 | 0.58 |
| *Planting phase fertilizer:* | | | | | | | |
| Organic | kg | 7,227 | 0.28 | 8,867 | 0.12 | 7,100 | 0.78 |
| *Flowering phase fertilizer:* | | | | | | | |
| Crystalline | kg | 12 | 1.83 | 39 | 0.92 | 24 | 0.96 |
| Artificial | kg | 55 | 1.85 | 0 | 0 | 41 | 1.56 |
| *Output* | | | | | | | |
| Yield | kg | 7,241 | 0.57 | 2,967 | 0.66 | 3,810 | 0.81 |

Note: SE, scale efficiency; DRS, decreasing returns to scale; IRS, increasing returns to scale; CV, coefficient of variation which is defined as follows: CV = $\frac{\sigma}{\mu}$ that is shown as a coefficient in this study with σ indicating the covariate's standard deviation and μ the covariate's mean value.

A greenhouse pepper farm facing DRS or diseconomies of scale may choose to decrease its size to reduce some of the excessive use of inputs. In the region of Peja, Mitrovica, and Gjilan, farmers interviewed reasoned that the high input prices often led them to a reduction of the inputs consumed. For the greenhouse pepper farms in the situation of IRS or economies of scale, however, an important decision was to achieve an overall higher use of inputs. In this frame of analysis, an increase in output may be achieved through an increase in the use of inputs where previously Table 2.9 noted the discrepancy of mean values among IRS relative to the SE greenhouse pepper farms. Figure 2.3 shows greenhouse pepper farms' efficiency frontiers with yield in kg as the output, and insecticide, labor, greenhouse area in $m^2$, and planting phase organic, flowering phase crystalline and flowering phase artificial fertilizers as the inputs. The depiction of the model is under VRS using the BCC model, and CRS using the CCR model with their compatible efficiency frontiers. They suggested many farms that were operating in the areas of IRS and DRS may become scale efficient through further input use improvements. A point on the piecewise linear curve representing the BCC model shows that a farm has received an efficiency score of 1.00, and the few points on the upward sloping line representing the CCR model indicate efficiency score values of 1.00 and with zero-slacks. In contrast to the greenhouse tomato efficiency frontiers, small-sized and large-sized farms appeared to have a lower disparity in the use of inputs with efficiency scores close to each other.

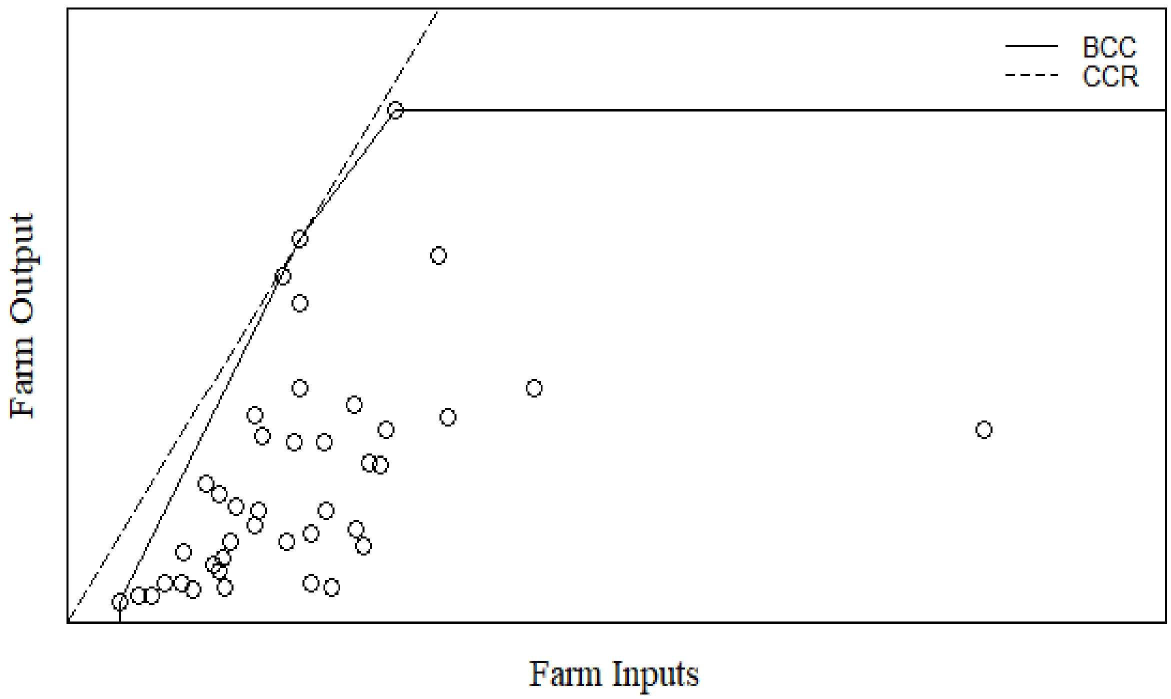


Figure 2.3. Greenhouse Pepper Farms' Efficiency Frontiers
The BCC model represents the VRS technology with the piecewise linear curve, and the CCR model represents the CRS technology with the upward sloping line. Note that the plot includes multiple inputs. Thus, aggregation occurred when attempting to plot all the BCC- or CCR-efficient farms on the frontiers using the Benchmarking package in R.

*Greenhouse Pepper Input Use Comparison at a Regional Level*

The aggregate efficiency scores with the VRS technology and BCC input-oriented efficiency found Prishtina the most efficient region with a mean efficiency of 0.99. Ferizaj followed with a mean efficiency of 0.93 in the use of inputs when producing greenhouse peppers. There is an increasing number of farmers from Prishtina receiving financial support by the Ministry of Agriculture, Forestry and Rural Development (MAFRD) as well as by the Municipality of Prishtina in the production of greenhouse peppers among other vegetable crops. This and the fact that Prishtina is the capital of Kosovo with a large market for the greenhouse peppers fosters the study's result that this region may have the most efficient producers in the use of inputs. Under the CRS technology and CCR input-oriented efficiency, region of Ferizaj with a mean efficiency of 0.87 ranked first, followed by Prishtina with a mean efficiency of 0.80.

The efficiency ranking of the regions under VRS was of relevance for consideration given the same reasons in the case of the greenhouse tomato input efficiency use at a regional level. It was important to note that CCR assumed all the greenhouse pepper farms operate at an optimal scale. This assumption may not hold in Kosovo's newly emerging greenhouse sector. For this reason, the results suggesting Prishtina as the region with the most efficient producers were more likely to be consistent with the input efficiency use comparison at a regional level. Table 2.10 shows the mean efficiency of each greenhouse tomato producing region under PTE, TE, and SE.

To provide an input efficiency use representation of the regions of Prishtina, Ferizaj, Gjilan, Gjakova, Peja, and Mitrovica, Figure 2.4 shows the aggregate mean efficiencies. Additional statistics include the median, lower quartile, and upper quartile with the possibility to understand how the regions rank and differ with one another in the optimal use of inputs. The greenhouse pepper producing regions with producers having low efficiency may be observed in the lower quartile compared to the producers having high efficiency in the upper quartile. However, this study emphasized the mean efficiency of the regions. This statistic may exhibit how well each region uses inputs in the production of the greenhouse peppers.

Table 2.10. Greenhouse Pepper BCC, CCR, and Scale Efficiency Results

| Aggregate efficiency scores | | | | Efficiency score range and no. of farms | | | |
|---|---|---|---|---|---|---|---|
| Region | PTE | TE | SE | Range | PTE | TE | SE |
| *Prishtina* | | | | *Prishtina* | | | |
| Mean | 0.99 | 0.80 | 0.80 | <0.40 | 0 | 1 | 1 |
| Min | 0.95 | 0.27 | 0.27 | 0.40-0.69 | 0 | 0 | 0 |
| Max | 1.00 | 1.00 | 1.00 | 0.70-0.99 | 1 | 3 | 3 |
| St. Dev. | 0.02 | 0.28 | 0.28 | 1.00 | 5 | 2 | 2 |
| *Ferizaj* | | | | *Ferizaj* | | | |
| Mean | 0.93 | 0.87 | 0.93 | <0.40 | 0 | 0 | 0 |
| Min | 0.70 | 0.48 | 0.62 | 0.40-0.69 | 0 | 3 | 1 |
| Max | 1.00 | 1.00 | 1.00 | 0.70-0.99 | 4 | 2 | 4 |
| St. Dev. | 0.11 | 0.18 | 0.13 | 1.00 | 8 | 7 | 7 |
| *Gjilan* | | | | *Gjilan* | | | |
| Mean | 0.84 | 0.36 | 0.43 | <0.40 | 0 | 2 | 2 |
| Min | 0.66 | 0.19 | 0.27 | 0.40-0.69 | 1 | 3 | 3 |
| Max | 1.00 | 0.46 | 0.59 | 0.70-0.99 | 3 | 0 | 0 |
| St. Dev. | 0.14 | 0.12 | 0.15 | 1.00 | 1 | 0 | 0 |
| *Gjakova* | | | | *Gjakova* | | | |
| Mean | 0.89 | 0.77 | 0.86 | <0.40 | 0 | 0 | 0 |
| Min | 0.72 | 0.57 | 0.69 | 0.40-0.69 | 0 | 3 | 1 |
| Max | 1.00 | 1.00 | 1.00 | 0.70-0.99 | 3 | 1 | 3 |
| St. Dev. | 0.12 | 0.19 | 0.13 | 1.00 | 3 | 2 | 2 |
| *Peja* | | | | *Peja* | | | |
| Mean | 0.86 | 0.56 | 0.66 | <0.40 | 0 | 3 | 1 |
| Min | 0.40 | 0.23 | 0.23 | 0.40-0.69 | 1 | 3 | 4 |
| Max | 1.00 | 0.96 | 0.96 | 0.70-0.99 | 5 | 3 | 4 |
| St. Dev. | 0.20 | 0.27 | 0.25 | 1.00 | 3 | 0 | 0 |
| *Mitrovica* | | | | *Mitrovica* | | | |
| Mean | 0.85 | 0.39 | 0.46 | <0.40 | 0 | 3 | 2 |
| Min | 0.58 | 0.22 | 0.27 | 0.40-0.69 | 1 | 1 | 2 |
| Max | 1.00 | 0.65 | 0.65 | 0.70-0.99 | 1 | 0 | 0 |
| St. Dev. | 0.20 | 0.18 | 0.17 | 1.00 | 2 | 0 | 0 |

Note: PTE, pure technical efficiency; TE, technical efficiency; SE, scale efficiency.

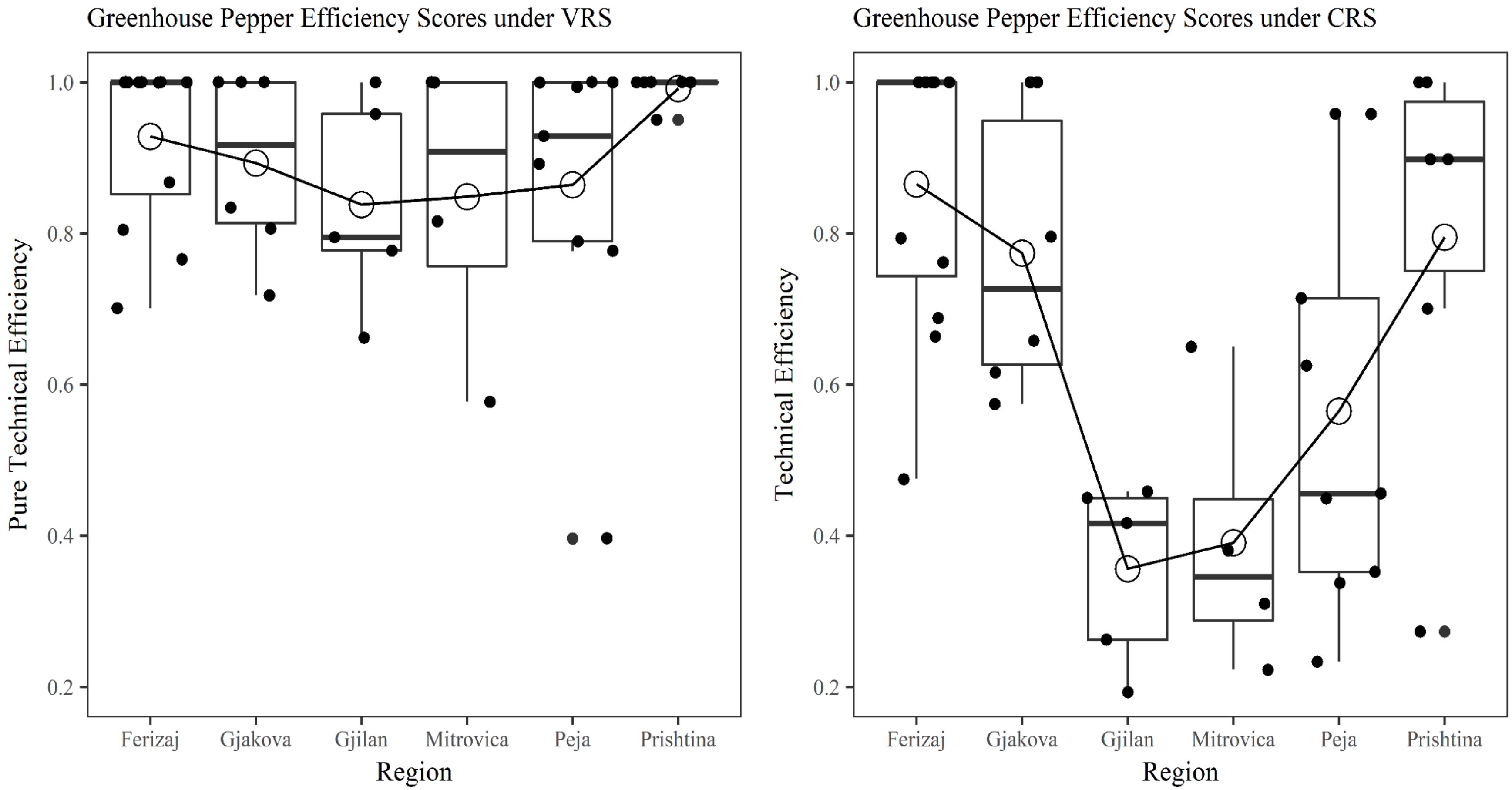


Figure 2.4. Efficiency Representation of Greenhouse Pepper Producing Regions.
Note: Circles connected by lines show each region's aggregate mean efficiency. Statistics presented are minimum, first quartile, median, mean, third quartile, and maximum. Under the VRS technology notice that region of Prishtina has a high input efficiency use, it has also a close to zero standard deviation, and the same median and maximum value. Therefore, the shape of the boxplot appears different.

*Linear and Logistic Regression Implications*

The external and production-related variables for the linear and logistic regression analyses were found to be different among greenhouse tomato and pepper farms. Some variables had significant impacts on the optimal use of inputs among greenhouse tomato farms and others among greenhouse pepper farms. As in the selection of the variables for the input-oriented DEA models, the Akaike information criterion (AIC), statistical significance of the variables and the literature review were considered to choose appropriate variable sets. As shown in Table 2.11, some variables included in the final models differed across tomatoes and peppers. The estimation procedure for the greenhouse tomato farms included the need to participate in a crop nutrition training, power source, rows per greenhouse, wholesale price per kg of tomatoes, irrigation equipment value in euros, education in years, and number of family members. While, the estimation procedure for the greenhouse pepper farms included an indicator variable for the farmer's external revenue, farmer market price per kg of peppers, other greenhouse crops grown, well depth in meters, irrigation equipment value in euros, education in years, and number of family members.

The logistic regression analysis for the greenhouse tomato farms presented in Table 2.11 found rows per greenhouse to have a positive and statistically significant impact on input efficiency use. The discrepancy in the number of tomato rows between a farmer having a single tunnel greenhouse and a farmer with a multi tunnel greenhouse noted in the study supports further the positive effect of the variable. This result may suggest that farmers having greenhouse rows occupied with crops for non-commercial uses may make them less efficient in the optimal use of inputs. The positive result may also suggest that crops grown for home consumption may limit the greenhouse area for the greenhouse tomatoes produced for the market. The variable electricity utilized as the power source compared to fuel had a negative and significant impact on input efficiency use. The negative and significant coefficient of the power source variable showed that farmers using electricity were less efficient in the use of inputs. Given that electricity remains a costly source of power in Kosovo (Bowen, Myers, Myderrizi, Hasaj, & Halili, 2013), this result was expected to demonstrate a negative impact on input efficiency use. In line with earlier expectations, electricity and the high costs associated with it may not promote an optimal use of inputs relative to the alternative of using fuel.

An important observation was the negative and significant coefficient of the farmer's need for a vegetable crop nutrition training in the production of greenhouse tomatoes. The survey's Likert scale question result that 83 percent of the farmers strongly agreed to participate in a crop nutrition training further supported the negative coefficient of the variable that some of the farmers were not using optimally the inputs. The group of farmers having strongly agreed to participate in a crop nutrition training may be the same group of farmers found operating under DRS. For example, a DRS farm was found overutilizing the planting phase organic and artificial fertilizers, and the flowering phase artificial and crystalline fertilizers relative to the scale efficient farms.

This study's result conformed to that of Coelli, Rahman, and Thirtle (2002) that education had a negative impact on efficiency, except that this study has a different context and more years of education had a negative and significant impact on greenhouse tomato input efficiency use. Another prevalent characteristic that greenhouse tomato farmers showed during the interview process was the high level of practical experience in production. Often the greenhouse farmer's perception was that the production expertise may be of more importance on the efficient use of inputs than more years of education. This was explained by the negative and statistically significant coefficient of the variable in the model. The remaining variables such as irrigation equipment value in euros (p-value = 0.15), wholesale price per kilogram (kg) of tomatoes (p-value = 0.15), and number of family members (p-value = 0.19) were found to be insignificant.

On the contrary, the logistic regression model for greenhouse pepper farms presented in Table 2.11 included external revenue, farmer market price per kg of peppers, other greenhouse crops grown, well depth in meters, irrigation equipment value in euros, education in years, and the number of family members. Estimates from this model show the number of family members to have a negative and statistically significant impact. Although a result with a positive effect was expected, the negative coefficient may be explained by the fact that the greenhouse pepper farms relative to the greenhouse tomato farms required less active working days in the management of farm operations. An additional family member to conduct the farms operations in the greenhouse may not necessarily lead to higher levels of input efficiency use. Well depth in meters was a variable that showed a positive and statistically significant impact on the efficient use of inputs. When faced with increased levels of water scarcity, farms in regions with deeper wells may be more able to irrigate during the flowering season. It may be possible that depth of the well was likely to affect the optimal use of inputs notably when using crystalline fertilizer. This result was consistent with the expectations of the study.

Education in years (p-value = 0.47), irrigation equipment value in euros (p-value = 0.38), other greenhouse crops grown (p-value = 0.34), external revenue (p-value = 0.21), and farmer market price per kg of peppers (p-value = 0.89) were found to be statistically insignificant. Table 2.11 provides the logistic regression results with the statistical significance of the variables and Figure 2.5 presents a ranking of the variables based on the variable’s positive and negative impact. For a more direct interpretation of the results ranking was performed using the odds ratios of the estimates. The rows per greenhouse variable showed a high positive effect and electricity an opposite effect on the input efficiency use among the greenhouse tomato farms. While, well depth in meters demonstrated a high positive effect and the number of family members a high negative effect in the optimal use of inputs among the greenhouse pepper farms.

Table 2.11. Logistic Regression Results

| | Greenhouse Tomato Model | | | | Greenhouse Pepper Model | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 95 % CI | | | | 95 % CI | |
| Variable | β (SE) | OR | Lower | Upper | β (SE) | OR | Lower | Upper |
| Crop nutrition training | -1.853** (0.917) | 0.157 | 0.024 | 0.952 | | | | |
| Power source (electricity or fuel) | -2.192*** (0.840) | 0.112 | 0.017 | 0.506 | | | | |
| Rows per greenhouse | 0.172** (0.070) | 1.188 | 1.059 | 1.390 | | | | |
| Wholesale price per kg | -4.694 (3.251) | 0.009 | 0.000 | 3.413 | | | | |
| External revenue | | | | | -1.856 (1.487) | 0.156 | 0.005 | 2.094 |
| Farmer market price per kg | | | | | 0.120 (0.876) | 1.127 | 0.197 | 7.300 |
| Other crops grown | | | | | -0.667 (0.702) | 0.513 | 0.115 | 1.991 |
| Well depth in meters | | | | | 0.517** (0.255) | 1.677 | 1.102 | 3.167 |
| Irrigation in euro value | -0.001 (0.001) | 0.999 | 0.996 | 1.000 | 0.003 (0.003) | 1.003 | 0.997 | 1.010 |
| Education in years | -0.336** (0.170) | 0.714 | 0.484 | 0.956 | 0.141 (0.197) | 1.152 | 0.785 | 1.768 |
| Family members | 0.092 (0.070) | 1.097 | 0.951 | 1.287 | -0.862** (0.411) | 0.422 | 0.159 | 0.841 |
| (Constant) | 2.384 (1.893) | | | | -1.086 (3.925) | | | |
| Observations | 94 | | | | 42 | | | |
| Log Likelihood | -25.320 | | | | -14.856 | | | |
| Akaike Inf. Crit. | 66.640 | | | | 45.712 | | | |
| Nagelkerke $R^2$ | 0.492 | | | | 0.523 | | | |

Note: β, regression coefficient; SE, standard error; OR, odds ratio; CI, confidence interval. The statistical significance of the variables is represented by *p<0.1; **p<0.05; ***p<0.01. There were other variables not shown in the logistic model above which have been tested and found insignificant. Such variables included the indicator variable grant and the grant value of the greenhouse tomato and pepper farmers who received grants with specific euro values from the Ministry of Agriculture, Forestry and Rural Development (MAFRD). In multiple model combinations in R programming, no set of variables was found using the dummy variable grant or the grant value in euros received by the farmer with a statistically significant impact in determining an efficient greenhouse farm. Considering these reasons, this study did not include any of the two variables in the logistic regression model.

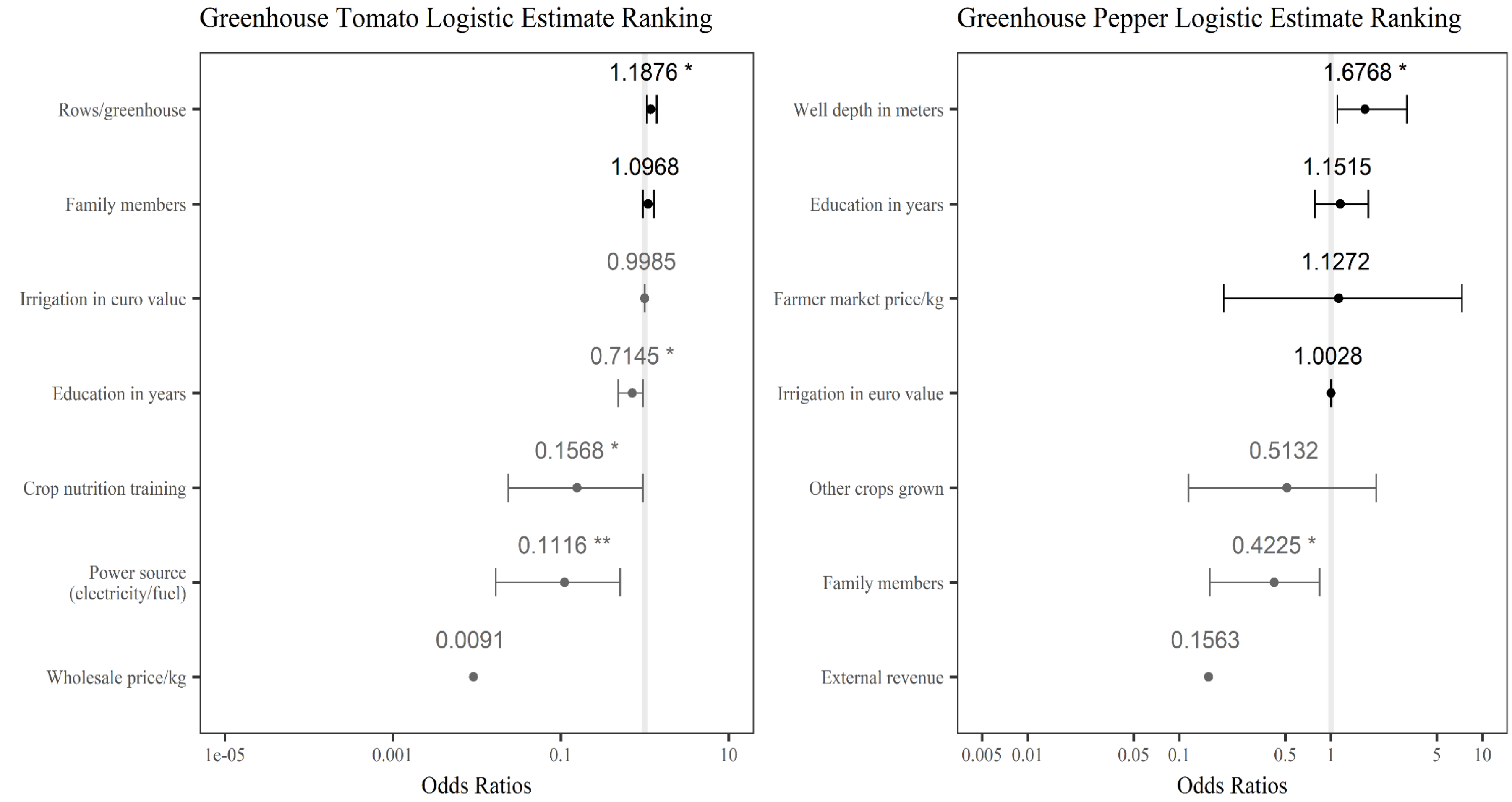


Figure 2.5. Logistic Regression Variable Ranking for Greenhouse Tomato and Pepper Farmers.
Note: The vertical line represents the zero-effect line, while the error bars show the 95 percent confidence interval. Variables with a low range of confidence intervals show small error bars. The position of the variables to the left of the zero-effect line exhibit negative effects on input efficiency use. The statistical significance of the variables is represented by *p<0.05; and, **p<0.01.

Comparatively, the linear model for the greenhouse tomato farms with TE scores as the scalar dependent variable presented in Table 2.12 found statistically significant rows per greenhouse with a positive coefficient. This variable was statistically significant at the one percent confidence interval compared to the five percent confidence interval in the logistic regression. The implication of the positive impact of the greenhouse tomato rows implied that some of the greenhouse tomato farmers may increase the number of rows with greenhouse tomatoes and potentially contribute to a further optimal use of inputs. In addition, the precise application of inputs was an issue farmers faced given that new technology was absent for a proportionate use of inputs in most farms. For example, none of the farmers interviewed reported having a heating system and only one farmer with an automatic ventilation system designed in the production of greenhouse tomatoes and peppers. More greenhouse tomato rows may lead to an understanding that in the absence of farm technologies, it may impact the optimal use of inputs among greenhouse tomato farms. In this line of reasoning, the result of power source with a significant negative coefficient suggested that electricity as the power source compared to fuel contributed to a less efficient use of the inputs among greenhouse tomato farms. In line with earlier expectations, the use of electricity may be perceived costlier compared to fuel in performing some of the greenhouse farm operations. The result noted that using overall electricity over fuel in the production of greenhouse tomatoes may not promote a more efficient use of inputs.

Variables reported statistically insignificant in the linear regression model included the need for a crop nutrition training, irrigation equipment value in euros, education in years, and the number of family members. The variable wholesale price per kg of tomatoes found insignificant with the logistic regression model was in fact significant and with a negative coefficient in the linear regression model. It was of crucial interest to reflect on this result given farmers' concern reported during the interview process that the price received per kg of tomatoes from the vegetable wholesalers in Kosovo was hurting their gross margins. The pressure originating from low profits and high input costs may lead farmers to lower efficient levels in the use of inputs. First, the negative and statistically significant coefficient of the wholesale price in the linear regression model supported further the assumption that selling to the vegetable wholesalers may result to a less efficient use of inputs. Second, as the purchase of costly inputs is often realized with difficulty among greenhouse tomato farms, this may encourage a lower and disproportionate application of inputs. For example, a farmer attempting to contract his production costs may choose to reduce the amount of inputs used without a comprehensive analysis of the effects generated in the production process.

For the greenhouse pepper farms, the model presented in Table 2.12 showed the number of family members statistically significant and negative. Despite that a result with a positive effect was expected, the linear regression model strengthened the assumption that a higher presence of the family members during the growing season may not contribute to an efficient use of inputs. This may be true among single tunnel greenhouses where labor may not be required as intensively as in the multiple tunnel greenhouses. The farmer market price per kg of peppers and more years of education were shown to be positive and insignificant in both models. However, other greenhouse crops grown and irrigation value in euros reported as insignificant in the logistic regression model were significant at the one and five percent confidence interval in the linear regression model, respectively. According to the linear regression model, farms that cultivated other vegetable crops with peppers in the same growing season and greenhouse were influenced unfavorably in the efficient use of inputs. The negative coefficient of growing other vegetable crops with peppers indicated that more vegetable crops grown with peppers may lead to a situation where it is unlikely to achieve an optimal use of inputs. The different crop nutrition ratio requirements and the lack of technology in the use of inputs for each crop in the greenhouse may be among the reasons leading to a negative and statistically significant coefficient.

Another variable that became significant with a positive coefficient under the linear regression model was the irrigation equipment value in euros. This may be due to the use of the TE scores as a scalar dependent variable in the linear regression model. The result indicated that irrigation equipment of higher value may increase the input efficiency use in the production of greenhouse peppers. In fact, inadequate irrigation of the vegetable crops because of the old irrigation equipment may constrain the input efficiency use. Despite the wide presence of the drip irrigation systems in Kosovo, this result further supported the statement of the Balliu and Kaçiu (2008) that the frequency and amount of the irrigation needs more improvement. The greenhouse pepper farmers during the data collection process reported irrigation equipment with varying euro values. Therefore, it may be from the model's estimation that irrigation equipment of higher euro value may result to a more efficient use of inputs.

A key finding deriving from the comparison of the logistic regression model to the linear regression model employing the same set of variables was that under the latter model more variables became statistically significant and with higher confidence intervals. However, the positive or negative direction of the coefficients was maintained in general from model to model. It was of substantial importance to find from this study that having TE scores used as the binary outcome and with a threshold to indicate an (in)efficient greenhouse produced comparatively different results to using TE scores as a scalar dependent variable. Overall, variables were more statistically significant in the linear regression models. There were variables that were progressing from the logistic regression model's five percent confidence interval to that of the linear regression model's one percent confidence interval.

This showed that using TE scores without a threshold and as a scalar dependent variable may lead to higher statistical significance.

Table 2.12 shows the model estimate specifics discussed for the evaluation of both greenhouse tomato and pepper input efficiency use and their corresponding statistical significance under the linear regression model. While, Figure 2.6 presents a ranking of the variables based on the estimate’s positive and negative impact. In contrast to the logistic regression model, the linear estimate ranking results included the beta estimate values of the variables. The variable including rows per greenhouse showed a high and positive effect and wholesale price per kg of tomatoes a high and opposite effect on the input efficiency use among the greenhouse tomato farms. While, the variable irrigation value in euros demonstrated a high and positive effect and other greenhouse crops grown a high and negative effect in the optimal use of inputs among the greenhouse pepper farms. The results showed that even in the ranking of the variables based on the positive, negative and statistically significant effects, the linear and logistic regression models produced comparatively different results.

Table 2.12. Linear Regression Results

| | Greenhouse Tomato Model | | | | Greenhouse Pepper Model | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 95 % CI | | | | 95 % CI | |
| Variable | β | SE | Lower | Upper | β | SE | Lower | Upper |
| Crop nutrition training | -0.144* | (0.074) | -0.288 | 0.001 | | | | |
| Power source (electricity or fuel) | -0.188*** | (0.053) | -0.292 | -0.084 | | | | |
| Rows per greenhouse | 0.012*** | (0.004) | 0.005 | 0.019 | | | | |
| Wholesale price per kg | -0.364** | (0.169) | -0.694 | -0.033 | | | | |
| External revenue | | | | | -0.015 | (0.083) | -0.177 | 0.148 |
| Farmer market price per kg | | | | | 0.044 | (0.062) | -0.077 | 0.166 |
| Other crops grown | | | | | -0.145*** | (0.050) | -0.244 | -0.047 |
| Well depth in meters | | | | | 0.023 | (0.014) | -0.004 | 0.050 |
| Irrigation in euro value | -0.0001 | (0.0001) | -0.0003 | 0.00003 | 0.0005** | (0.0002) | 0.0001 | 0.0009 |
| Education in years | -0.017* | (0.009) | -0.034 | -0.0001 | 0.007 | (0.014) | -0.021 | 0.034 |
| Family members | 0.005 | (0.007) | -0.007 | 0.018 | -0.054*** | (0.017) | -0.087 | -0.020 |
| (Constant) | 0.802*** | (0.145) | | | 0.905*** | (0.290) | | |
| Observations | 94 | | | | 42 | | | |
| $R^2$ | 0.309 | | | | 0.462 | | | |
| Adjusted $R^2$ | 0.252 | | | | 0.351 | | | |
| Residual Std. Error | 0.248 (df = 86) | | | | 0.224 (df = 34) | | | |
| F Statistic | 5.484*** (df = 7; 86) | | | | 4.173*** (df = 7; 34) | | | |

Note: β, regression coefficient; SE, standard error; CI, confidence interval. The statistical significance of the variables is represented by *p<0.1; **p<0.05; ***p<0.01. Like with logistic regression, the variables grant and the grant value in euros of the greenhouse tomato and pepper farmers who received grants with specific euro values from the Ministry of Agriculture, Forestry and Rural Development (MAFRD) were found to be insignificant. In many model combinations in R programming, no set of variables was found using the indicator or dummy variable grant or the grant value in euros with a statistically significant impact in the scalar dependent variable consisting of technical efficiency (TE) scores. Therefore, this study did not include any of the two variables in the estimation of the linear regression model shown above.

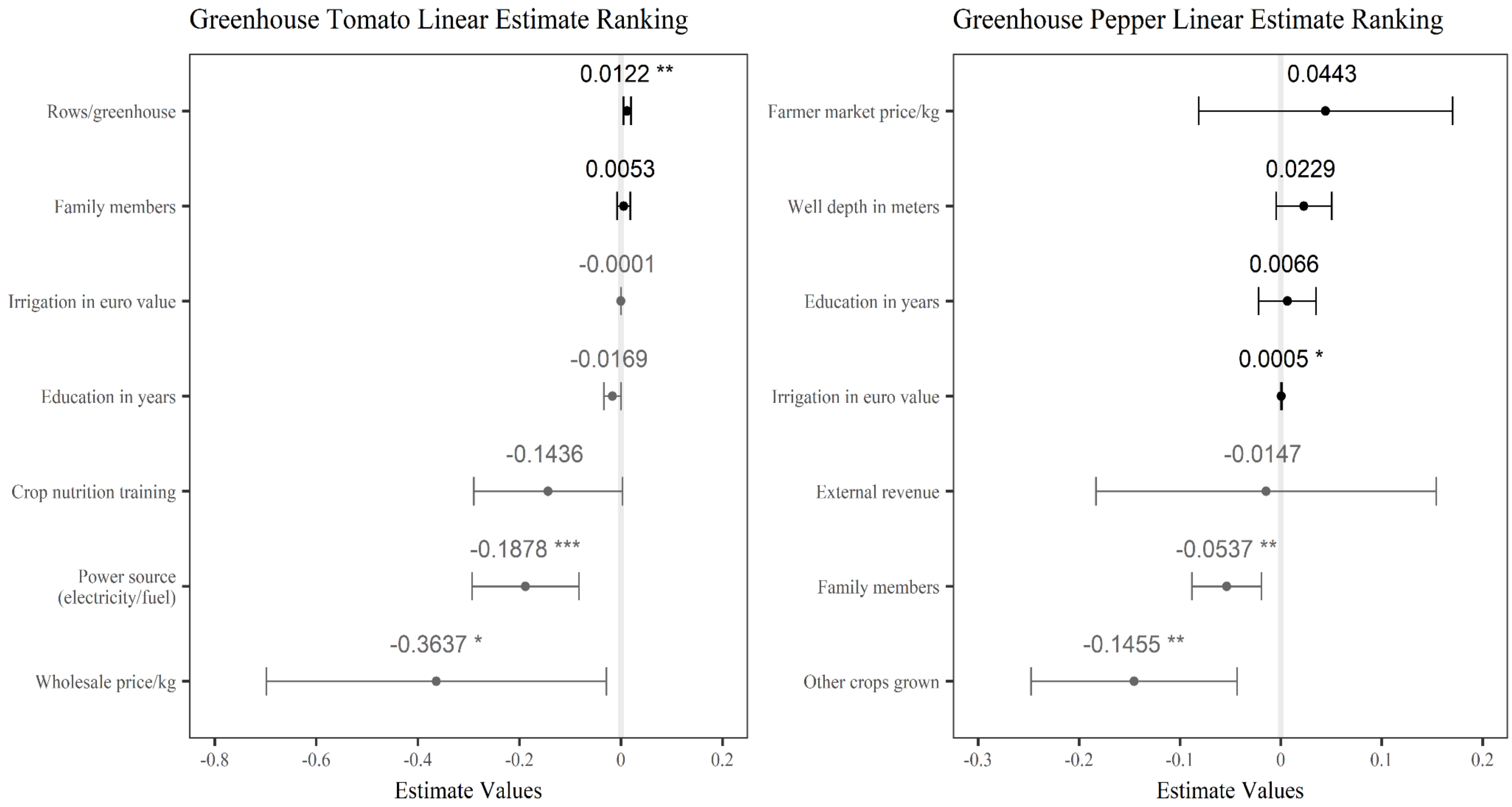


Figure 2.6. Linear Regression Variable Ranking for Greenhouse Tomato and Pepper Farmers.
Note: The vertical line represents the zero-effect line, while the error bars show the 95 percent confidence interval. Variables with a low range of confidence intervals show small error bars. The position of the variables to the left of the zero-effect line exhibit negative effects on input efficiency use. The statistical significance of the variables is represented by *p<0.05; and, **p<0.01.

## Conclusion

The study aimed to analyze greenhouse tomato and pepper input efficiency use in Kosovo at a farm and regional level and to determine the external factors that affect efficiency by using linear and logistic regression as a two-stage procedure to the data envelopment analysis (DEA) method.

In a regional analysis, the BCC model found the region of Prizren the most efficient in the use of inputs regarding the production of greenhouse tomatoes, and region of Prishtina in the production of greenhouse peppers, respectively. This can provide helpful insights for the greenhouse vegetable sector in Kosovo. The inefficient greenhouse producers found operating under decreasing returns to scale (DRS) may become scale efficient by reducing the use of agricultural inputs. Those operating under increasing returns to scale (IRS) may become scale efficient by increasing the use of agricultural inputs. Results suggest that for greenhouse tomato producing regions to achieve proper technical efficiency (TE), input use need to improve by 13 to 41 percent, depending on region. Comparatively, greenhouse pepper producing regions had more efficient producers. Therefore, potential improvements in efficient input use ranged across regions from 1 to 16 percent.

At a farm level, the BCC model showed that 33 percent of the greenhouse tomato farms were fully efficient and only 16 percent under the CCR model. While only 52 percent of greenhouse pepper farms were fully efficient under the BCC model and 26 percent under the CCR model, respectively. Given differences in scale size, there were farms that had complete optimal use of inputs. However, many of the greenhouse tomato farms under DRS and greenhouse pepper farms under IRS and DRS were struggling to find an optimal use of inputs. The results suggest also a policy is of vital interest to address the issue of selling greenhouse tomatoes with a price that may jeopardize the financial health and future of the farms. Region of Prizren with the most concentration of greenhouse tomato farms and region of Gjakova were found to be particularly influenced by the price received per kilogram (kg) of tomatoes from the vegetable wholesalers. Meanwhile, no estimates were statistically significant in relation to the price received by greenhouse pepper farms per kg of peppers relative to the price received per kg of tomatoes.

Totally, this study explores the inefficient input use in the production of greenhouse tomatoes and peppers, which is caused by two primary factors. One factor is the disproportionate use of inputs without a corresponding increase in yields noted among DRS farms. Often this leads to a loss of production which may be avoided through reducing the use of inputs to the same level of the scale efficient input values. The other factor is concerning the disadvantageous market conditions, where pressure from imports and low prices set from vegetable wholesalers heavily affect greenhouse tomato production. This study revealed that overall under the given greenhouse tomato and pepper production levels, there would be a large opportunity for the technically inefficient farms and regions to improve their whole performance in the use of inputs.

## References


Adhikaria, C. B., and T. Bjorndalb. 2012. "Analyses of Technical Efficiency Using SDF and DEA Models: Evidence from Nepalese Agriculture." *Applied Economics* 3297-3308.

Ahearn, M., and J. Sterns. 2013. "Direct-to-Consumer Sales of Farm Products: Producers and Supply Chains in the Southeast." *Journal of Agricultural and Applied Economics* 497-508.

Alboghdady, M. A. 2014. "Nonparametric Model for Measuring Impact of Inputs Density on Egyptian Tomato Production Efficiency." *International Journal of Food and Agricultural Economics* 81-90.

Andersson, R., H. Rexhepi, M. Farinelli, and J. D'Costa. 2001. *Statistics on Agriculture in Kosovo.* Pristina: Statistical Office of Kosovo.

Arcotrass Gmbh, Vakakis International SA, Eurocare Gmbh, AKI. 2006. *Kosovo Report.* Arcotrass-Consortium.

Balcombe, K., I. Fraser, M. Rahman, and L. Smith. 2006. "Examining the Technical Efficiency of Rice Producers in Bangladesh." *Journal of International Development* 1-16.

Balliu, A., and S. Kaçiu. 2008. *Potential of the Greenhouse Industry in Kosovo.* Horticultural Promotion in Kosovo (HPK).

Banker, R. D., W. W. Cooper, and A. Charnes. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 1078-1092.

Banker, R. D., W. W. Cooper, L. M. Seiford, R. M. Thrall, and J. Zhu. 2004. "Returns to Scale in Different DEA Models." *European Journal of Operational Research* 345-362.

Barragan, J., L. Cots, J. Monserrat, R. Lopez, and I. P. Wu. 2010. "Water Distribution Uniformity and Scheduling in Micro-irrigation Systems for Water Saving and Environmental Protection." *Biosystems Engineering* 202-211.

Barry, P. J., J. A. Hopkin, and C. B. Baker. 1988. *Financial management in agriculture*. Interstate Printers and Publishers.

Battilani, P., A. Pietri, C. Barbano, A. Scandolara, T. Bertuzzi, and A. Marocco. 2008. "Logistic Regression Modeling of Cropping Systems to Predict Fumonisin Contamination in Maize." *Journal of Agricultural and Food Chemistry* 10433–10438.

Blank, S. C., and K. W. Erickson. 2007. "Agricultural Household Hedging with Off-Farm Income." *Western Economics Forum*.

Błażejczyk-Majka, L., R. Kala, and K. Maciejewski. 2012. "Productivity and Efficiency of Large and Small Field Crop Farms and Mixed Farms of the Old and New EU Regions." *Agricultural Economics (Zeme de lská Ekonomika)* 61-71.

Bowen, B. H., J. A. Myers, A. Myderrizi, B. Hasaj, and B. Halili. 2013. Kosovo Household Energy Consumption Facts and Figures.

Caswell, M. F., and D. Zilberman. 1986. “The Effects of Well Depth and Land Quality on the Choice of Irrigation Technology.” *American Journal of Agricultural Economics* 798-811.

Charnes, A., W. Cooper, and E. Rhodes. 1978. “Measuring the Efficiency of Decision Making Units.” *European Journal of Operational Research* 429-444.

Chivuraise, C., T. Chamboko, and G. Chagwiza. 2016. “An Assessment of Factors Influencing Forest Harvesting in Smallholder Tobacco Production in Hurungwe District, Zimbabwe: An Application of Binary Logistic Regression Model.” *Advances in Agriculture*.

Coelli, T., S. Rahman, and C. Thirtle. 2002. “Technical, Allocative, Cost and Scale Efficiencies in Bangladesh Rice Cultivation: A Non-parametric Approach.” *Journal of Agricultural Economics* 607-626.

Cooper, W. W., L. M. Seiford, and K. Tone. 2002. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Kluwer Academic Publishers.

Dumas, Y., M. Dadomo, G. Di Lucca, and P. Grolier. 2003. “Effects of Environmental Factors and Agricultural Techniques on Antioxidantcontent of Tomatoes.” *Journal of Science, Food and Agriculture* 369-382.

El Benni, N., R. Finger, and S. Mann. 2012. “The Effect of Agricultural Policy Change on Income Risk.” *123rd EAAE Seminar*.

United Nations, Food and Agriculture Organization. 2017. *Good Agricultural Practices for Greenhouse Vegetable Production in the South East European Countries: Principles for Sustainable Intensification of Smallholder Farms*.

Haile, M. G., M. Kalkuhl, and M. A. Usman. 2015. “Market Information and Smallholder Farmer Price Expectations.” *African Journal of Agricultural and Resource Economics* 297-311.

Heeb, A., B. Lundegårdh, G. Savage, and T. Ericsson. 2006. “Impact of Organic and Inorganic Fertilizers on Yield, Taste, and Nutritional Quality of Tomatoes.” *Journal of Plant Nutrition and Soil Science* 535-541.

Huang, C. L., and B. H. Lin. 2007. “A Hedonic Analysis of Fresh Tomato Prices among Regional Markets.” *Review of Agricultural Economics* 783-800.

Huat, J., T. Doré, and C. Aubry. 2013. “Limiting Factors for Yields of Field Tomatoes Grown by Smallholders in Tropical Regions.” *Crop Protection* 120-127.

Jaleta, M., and C. Gardebroek. 2007. “Farm-gate Tomato Price Negotiations under Asymmetric Information.” *Agricultural Economics* 245-251.

Judah, T. 2008. *Kosovo: What Everyone Needs to Know.* Oxford University Press.

Kaciu, S. 2008. “Study of the Current Situation of Pepper Production in Kosovo.”

Kaciu, S., I. Babaj, S. Aliu, and I. Demaj. 2016. “Potential of Protected Vegetable Production in Kosovo and Future Perspectives.” *Acta Hortic* 461-466.

Korhonen, P., and T. Joro. 2015. *Extension of Data Envelopment Analysis with Preference Information: Value Efficiency.* Springer.

Li, L., W. Wu, P. Giller, J. O’Halloran, L. Liang, P. Peng, and G. Zhao. 2018. “Open Accesslife Cycle Assessment of a Highly Diverse Vegetable Multi-cropping System in Fengqiu County, China.” *Sustainability* 10(4).

Manevska-Tasevska, G. 2013. “Farmers' Knowledge Attributes Contribute to Attaining Higher Farm Technical Efficiency: A Transition Economy Case.” *The Journal of Agricultural Education and Extension* 7-19.

Mader, P., A. FlieBbach, D. Dubois, L. Gunst, F. Padrout, and N. Urs. 2002. “Soil Fertility and Biodiversity in Organic Farming.” *Science* 1694-7.

Ministry of Agriculture, Forestry and Rural Development. 2016. *Green Report*.

Ministry of Economic Development. 2017. *Energy Strategy of the Republic of Kosovo 2017-* 2026.

Nikolla, M., M. Meco, J. Bou Dib, M. Belegu, I. Qinami, X. Dulja, and E. Kadiu. 2013. “Increasing the Efficiency of the Albanian Agricultural Farms Using the DEA Model.” *Journal of Food, Agriculture and Environment* 1286-1290.

Provenzano, G., L. Cots, J. Monserrat, D. Autovino, and J. Barragán. 2016. “Determining Optimal Seasonal Irrigation Depth Based on Field Irrigation Uniformity and Economic Evaluations: Application for Onion Crop.” *Journal of Irrigation and Drainage Engineering* 142(10).

Raheli, H., R. M. Rezaei, M. R. Jadidi, and H. G. Mobtaker. 2017. “A Two-stage DEA Model to Evaluate Sustainability and Energy Efficiency of Tomato Production.” *Information Processing in Agriculture* 342-350.

Rodriguez, J. C., N. L. Shaw, and D. J. Cantliffe. 2007. “Influence of Plant Density on Yield and Fruit Quality of Greenhouse-grown Galia Muskmelons.” *HortTechnology* 17(4).

Sauer, J., S. Davidova, and L. Latruffe. 2012. “Determinants of Smallholders’ Decisions to Leave Land Fallow: The Case of Kosovo.” *Journal of Agricultural Economics* 119–141.

Sephton, P. S. 2010. Spatial Arbitrage in Sarawak Pepper Prices. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 405-416.

Sohail, L. 2008. “Effect of Organic Fertilizer on Plant Performance in Greenhouse Vegetable Production.” MS thesis, University of Windsor.

Tanfani, E., and A. Testi. 2012. *Advanced Decision Making Methods Applied to Health Care.* Springer Science and Business Media.

Tomek, W. G., and H. M. Kaiser. 2014. *Agricultural Product Prices*. Ithaca: Cornell University Press.

Theodoridis, A. M., and M. M. Anwar. 2011. "A Comparison of DEA and SFA Methods: A Case Study of Farm Households in Bangladesh." *The Journal of Developing Areas* 95-110.

Thimmareddy, K., B. Desai, and S. Vinoda Kumar. 2013. "Uptake of NPK, Availability of NPK and Quality Parameters of Bt Cotton ('Gossypium hirsutum' L.) as Influenced by Different Bio-fertilizers and 'In-situ' Green Manuring under Irrigation." *International Journal of Agriculture, Environment and Biotechnology* 623-628.

Zalkuw, J., R. Singh, R. Pardhi, and A. Gangwar. 2014. "Analysis of Technical Efficiency of Tomato Production in Adamawa State, Nigeria." *International Journal of Agriculture, Environment and Biotechnology* 645-650.

Zuzaku, A. 2014. "Marketing Research is Path Towards the Development of Agriculture in Kosovo." *Economic and Social Development: Book of Proceedings. Publisher: Varazdin Development and Entrepreneurship Agency* 488-496.

## Chapter 3. Assessing Government Grants: Evidence from Greenhouse Tomato and Pepper Farmers in Kosovo


### Abstract

The study applies propensity score matching with logistic regression as the distance measure and genetic matching to evaluate the effects of the Ministry of Agriculture, Forestry and Rural Development (MAFRD) grant programs. The primary contributions of this study were to broaden the understanding about greenhouse tomato and pepper farmers' probability to receive a MAFRD grant, whether age, education in years, yield in kilogram (kg), greenhouse value in euros, greenhouse area in square meters ($m^2$), distance to the market in kilometers (km), and other greenhouse crops grown influence farmers' participation in a grant program, and whether grants have an impact on the farmers' seasonal income. The findings showed that farmers with higher prior greenhouse tomato or pepper yields and small total greenhouse areas in $m^2$ were more likely to participate in a grant program. The analysis revealed that the greenhouse tomato grantees relative to the non-grantees make 1,777 euros more per growing season. This result was positive and significant under genetic matching. For the greenhouse tomato and pepper subjects under propensity score matching using logistic regression as the distance measure, differences in income per growing season were positive, however, insignificant.

## Introduction

Development of Kosovo's agriculture has long been minimal due to the Kosovo war which erupted in 1998 and lasted until 1999, a conflict that shaped the dynamics of the agriculture and left it vulnerable with a loss of productivity. Later, Kosovo declared its independence in 2008 and became a young country with the objective of building social and economic institutions (International Monetary Fund, 2011).

After the war, government of Kosovo spent vast amounts of money to resume a well-functioning economy given that a wide loss of farm data was registered leading to a situation where it was not possible to verify the correct information on crop yields (Kosovo Report, 2006). A 2016 planned government budget for a rural development program with a value of 23 million euros has been initiated in Kosovo with the objective of improving the productivity of the crops and the quality of the agricultural products (Kerolli-Mustafa & Gjokaj, 2016). Agriculture with its greenhouse vegetable subsector was among the recipients of the government's strategic investment. Over the last decade, there were investments through grant schemes that went to support the greenhouse farms. Tomatoes and peppers may be among the main crops grown in greenhouses throughout Kosovo. Comparatively, tomatoes are more common than peppers (Kaciu, 2008), however, the Ministry of Agriculture, Forestry and Rural Development or MAFRD (2016) suggested that the cultivation of both tomatoes and peppers have marked significant increases.

Despite the promise and potential of the greenhouse tomato and pepper production, the evidence supporting how MAFRD has promoted the increase in production of both vegetable crops is mixed. In addition, more evidence from studies that examine the effect of MAFRD's grants on the production of greenhouse tomatoes and peppers is necessary given their wide market prevalence throughout Kosovo. Reliable evidence regarding the relationship between the government financial support and a farmer's higher income per growing season from the greenhouse tomato or pepper production may be valuable to policy makers. A comprehensive research-based analysis of the Kosovo's MAFRD grant program effects on the greenhouse vegetable subsector is vital to identify any positive influence among the greenhouse farms.

The agricultural production has intensified given the increasing support to the sector through grant programs (Miftari, Hoxha, & Gjokaj, 2016). The funding initiative of MAFRD titled "Measure 101: Investments in physical assets and agricultural households" offering new greenhouses, support for open-field production and/or storage warehouses for vegetable farmers has favored an expansion in the greenhouse production (MAFRD, 2016). This funding had a component of providing grants to facilitate farmers' efforts to purchase new greenhouses. The gradual restructuring of the greenhouse production sector in Kosovo often requires among farmers an upgrade in farm facilities and MAFRD grant programs have attempted to address this issue of importance.

Following the government's funding initiatives, it is of interest to evaluate their impact on specific greenhouse farms. This study considers if the recipients relative to the non-recipients of MAFRD grants for the purchase of new greenhouses designed to produce tomatoes and peppers have indicatively different seasonal incomes. There are two reasons why this study chose to research and examine the tomato and pepper crops. First, there is an increase in the greenhouse vegetable production, and tomatoes and peppers remain among the main vegetables produced (Kaciu, Babaj, Aliu, & Demaj, 2016). This study aims to conduct an empirical analysis to understand the background of the rise in greenhouse vegetable production and to explore more specifically the impact of the grants in favoring an expansion in greenhouse tomato and pepper production. Second, a further research analysis includes a comparison of the production differences between recipients and non-recipients of grants and the grants' impact on the seasonal income of the farmers. This is important given an increasing productivity inequality between small and large greenhouse tomato and pepper farms leading potentially to different farm income levels. The question of whether and to what extent seasonal income differences exist between grantees and non-grantees is examined in this study.

Since the early 2000s, no studies have looked at the financial determinants of the greenhouse tomato and pepper production in Kosovo. Following the early counterfactual framework (Rosenbaum & Rubin, 1983; Rubin, 1977), this study applies propensity score matching (PSM) with logistic regression as the distance measure and genetic matching to estimate casual treatment effects of the farmers who received and who did not a grant. The analysis including PSM with logistic regression as the distance measure and genetic matching allows to quantify the treatment effects of grants on the farmers' probability to participate in a grant program. An outcome analysis including greenhouse seasonal income differences is also supported by the methods.

The estimation procedure in both methods observes how participation in a grant program may be influenced by the covariates age, education in years, yield in kilogram (kg), greenhouse value in euros, greenhouse area in square meters ($m^2$), distance to the market in kilometers (km), and other greenhouse crops grown. A further literature review suggests several important papers which have reviewed feasible aspects of the propensity score matching methods (see e.g. D'Agostino, 1998; Terza, Basu, & Rathouz, 2008; Caliendo & Kopeinig, 2008).

**Purpose**

Kosovar scholars have conducted little research regarding farmer income and greenhouse size. Furthermore, the existing literature confirms an absence of empirical studies on the effects of Kosovo's government investment efforts to advance the greenhouse vegetable subsector, while acknowledging that there may be reports which have discussed the effectiveness of the Ministry of Agriculture, Forestry and Rural Development (MAFRD) grant programs (MAFRD, 2016; MAFRD, 2017). However, without an empirical assessment, any conclusions reached about the effectiveness of MAFRD's provision of grants for the greenhouse farmers may be misleading.

Relatedly, it is of interest to know what covariates and to what degree these covariates impact greenhouse tomato and pepper farmers to apply for investment grants. One important approach to understand the covariate differences between farmer grantees and non-grantees is the use of matching for the treatment and control farmer groups. There are many methods available to perform matching. One method may be favored over the other based on the objective of the researcher. However, no consensus has suggested in the literature the best matching method (Stuart, 2010; Ruiz, Stout, & Herlihy, 2017).

To evaluate the impacts of these grants, this study considers the use of two methods for the matching of MAFRD grantees and non-grantees. First, propensity score matching is performed using logistic regression as the distance measure with replacement and a matching ratio of 3:1 including the average treatment effect on the treated (ATT). To compare the results, genetic matching is used with replacement and the same matching ratio of 3:1 including ATT. Both matching methods include the analysis of determining the difference in seasonal income between greenhouse tomato and pepper grantees and non-grantees.

**Methods**

With the use of propensity score matching (PSM) with logistic regression as the distance measure and genetic matching, the average treatment effect on the treated (ATT) was of interest in the analysis of the greenhouse tomato and pepper average differences. ATT may be considered as the average causal effect (Ho, Imai, King, & Stuart, 2007), where this study was interested to understand its effect on farmers receiving grants from MAFRD.

Under PSM based on a logistic regression estimation, ATT was performed with a 3:1 matching ratio and replacement given the small number of greenhouse tomato and pepper grantees compared to non-grantees. There may be varying perspectives whether PSM could be used with a relatively small sample size (Dehejia & Wahba, 2002; Zhong, 2004; Jacovidis, 2017), and the literature may not be clear on the sample size and on the proper comparison to treatment ratios (Jacovidis, 2017). However, this study relied in part on the previous research which used a matching ratio of 3:1 for PSM analysis (see e.g. Tabak, Zilberberg, Johannes, Sun, & McDonald, 2013; Birkbak, et al., 2016), and on the fact that in this study's sample there were few grantees relative to non-grantees. Therefore, to ensure matching of the treatment and control groups of farmers, a matching ratio of 3:1 was of relevance. Following that, an additional procedure was using genetic matching with replacement and the same matching ratio of 3:1. A comparison of the results produced by the two methods was used to examine any effect of the MAFRD grant programs. Moreover, the result comparison may enable to explore any influence on the greenhouse tomato and pepper farmers' participation in a grant program. And on the farmers' seasonal incomes as a potential pre-specified outcome of the matching methods.

*Propensity Score Matching with Logistic Regression*

This study employing the first method estimated the propensity scores using the binary logistic regression. The binary treatment case estimated the probability of participation versus nonparticipation (Caliendo & Kopeinig, 2008) or the probability of whether they received a grant or not with our dataset of greenhouse tomato and pepper farms. The equation of the binary logistic regression was as follows.

$$\ln\left(\frac{P_i(T_i = 1)}{1 - P_i(T_i = 1)}\right) = \widehat{\beta_0} + \widehat{\beta_1} X_{1i} + \ldots + \widehat{\beta}_n X_{ni} + e_i \tag{1}$$

This equation calculated the probability of a greenhouse farmer being in the group of grantees and divided it by the probability of being in the group of non-grantees. Where, $T_i = 0$ was for non-grantees, $T_i = 1$ for grantees, $X_{1i} \ldots X_{ni}$, covariates corresponded to the i[th] subject, and $e_i$ was the random error. Fitted values from equation (1) were used with the propensity score bounded between zero and one in the following equation. A greenhouse tomato or pepper grantee was defined as $T_i = 1$ and $T_i = 0$ otherwise. Covariates represented by $X_i$ were projected to have an impact on $T_i$, and Pr indicated the probability of the treatment group. Details of the equation were explained in Rosenbaum and Rubin (1983) and Joffe and Rosenbaum (1999).

$$p(X_i) = \Pr(T_i = 1|X_i) \tag{2}$$

In conjunction with the method of PSM employing logistic regression as the distance measure and the next matching method of genetic matching, the average treatment effect on the treated (ATT) played a significant role in explaining the average differences. The mathematical equation of ATT was presented in Ho, Imai, King, and Stuart (2007) and defined as follows.

$$\text{ATT} \equiv \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} T_i E\left[Y_i(1) - Y_i(0)|X_i\right] = \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} T_i \left[\mu_1(X_i) - \mu_0(X_i)\right] \tag{3}$$

The equation ATT estimated the average treatment effect for the greenhouse tomato and pepper grantees. The estimation procedure included $T_i = 1$ as the treatment group with $X_i$ covariates for the i[th] subject, and $Y_i = 1$ was the expected outcome for grantees and $Y_i = 0$ for non-grantees.

*Genetic Matching*

The practical use of the genetic matching algorithm corresponded to the research interest of the study notably in providing a second and automatic method to compare the results between the treatment group subjects and the control group subjects. Prior to the discussion of the genetic matching, it is useful to present Mahalanobis as a common distance measure in PSM methods. This distance measure was developed before PSM (Cochran & Rubin, 1973). In fact, it is not per se a matching method but more as a distance measure used together with the matching techniques. The equation of Mahalanobis is explained in detail prior to the elaboration of the genetic matching by Sekhon (2011), where S indicated the sample covariance matrix of X, and is presented as follows.

$$md(X_i,X_j) = \{(X_i-X_j)^T S^{-1} (X_i-X_j)\}^{1/2} \tag{4}$$

Genetic matching searched for the possibility to find a measure pertinent to achieve covariate balance after matching. The equation of this algorithm is obtained from Diamond and Sekhon (2013), who showed that genetic matching is performed by reducing a generalized version of the Mahalanobis distance (GMD). In contrast to the Mahalanobis distance, genetic matching includes an extra weight parameter W.

$$GMD(X_i,X_j,W) = \sqrt{(X_i-X_j)^T (S^{-1/2})^T \ WS^{-1/2}(X_i-X_j)} \tag{5}$$

This matching algorithm has been also discussed in detail (see e.g. Mebane & Sekhon, 1998). An additional procedure was to allow control subjects be used as matches for more than a treated unit on a matching ratio of 3:1. In this study's dataset, there was four times more non-grantees relative to the grantees producing greenhouse tomatoes and peppers. To ensure that a farmer who received a grant (treatment group) has a proper match with a non-grantee (control group), this study considered the use of replacement. In the literature it is noted that matching with replacement arguably provides better matches (Stuart & Rubin, 2008), and is preferred to use when possible in methods with a control group that may have similar values relative to a treatment group (Dehejia & Wahba, 2002).

*Covariate Theoretical Consideration*

Age, yield in kg, education in years, greenhouse value in euros, greenhouse area in square meters ($m^2$), distance to the market in kilometers (km), and other greenhouse crops grown were all covariates initially included in the model. It is recommended to use the literature review as a basis to identify a relevant covariate set (Luellen, Shadish, & Clark, 2005; Howarter, 2015). In addition, farmers have found the purchase of fertilizers, pesticides, and other farm supplies difficult in the production of greenhouse tomatoes and peppers in Kosovo. This may lead this study to include the farmer's seasonal income as the pre-specified outcome variable. Income levels may be among the main indicators as how well greenhouse tomato and pepper farmers perform in the vegetable sector.

The first covariate that this study may choose is the farmer's age. Literature suggested that age may be used as a prospective covariate in the propensity score matching methods (Howarter, 2015). There were many studies including age in models for the estimation of propensity scores (see e.g. Ali, Sharif, Mahmood, & Akmal, 2013; Wang, Xin, Li, & Yan, 2016). Using this covariate, this study's research may observe whether grantees and non-grantees with differing ages may be matched while considering their greenhouse tomato and pepper production specifities. First, it may help to discern if age is a factor in determining who participates in the grant program of the Ministry of Agriculture, Forestry and Rural Development (MAFRD). Second, age may seem relevant to observe whether grantees or non-grantees make more income over the course of a growing season. Examining education in years may be a covariate that can determine a farmer's participation in a MAFRD grant program. Using propensity score matching, education was found positive and significant for cherry production (Ali, Sharif, Mahmood, & Akmal, 2013). Education was found to be an important factor affecting a farmer's income (Panda, 2015), and wealth accumulation (Mahmudul, 2016). Moreover, agricultural education over practical experience of the farm manager can significantly influence productivity gains (Fîntîneru & Madsen, 2012). In fact, this study hypothesized that education may impact the farmer's income particularly among MAFRD grantees and non-grantees.

The covariate yield in kg has been considered an indicator of productivity and often of success among farms. It may be of interest to know how the farmer's yields influence the probability of participation in a grant program. In the developing countries, yield increases may be due to the use of nitrogen fertilizer, varieties and chemicals (Jaggard, Qi, & Ober, 2010). There may be a tendency of yields to stagnate which in fact may have previously improved (Ray, Ramankutty, Mueller, West, & Foley, 2012). The literature is mixed on the yield increases and decreases and factors that impact them. However, it may be expected that higher yields can have a positive effect on the farmer's participation in a MAFRD grant program. It was also hypothesized that the value of the greenhouse has an impact in the probability of receiving a grant. Some farmers having greenhouse structures with upgraded designs may suggest different production levels relative to farmers with traditional greenhouses. The volume of production may vary with the greenhouse value. In discussions with experts in Kosovo, it was understood that a greenhouse farmer with a relatively old greenhouse covering an area of 500 $m^2$ and may not produce as much as a greenhouse farmer with a new greenhouse covering an area of 500 $m^2$. The greenhouse environment may heavily impact crop cultivation, and the right climate growth conditions within the greenhouse can dictate the efficiency of the plant production (Yang & Simbeye, 2013). In Kosovo, greenhouses may lack proper ventilation systems. While, automatic ventilation can be effective in managing temperatures inside the greenhouse for high productivity (Kwon, et al., 2006). In fact, this study expects a positive and significant effect of the covariate greenhouse value in euros on a farmer's probability to participate in a MAFRD grant program.

The covariate greenhouse area in square meters ($m^2$) may enable a test of whether small-sized greenhouses are more likely to participate in a grant program compared to the large-sized greenhouses. Depending on the initial size of the farms, small farms can grow faster than large farms (Shapiro, Bollman, & Ehrensaft, 1987; Akimowicz, et al., 2013), however, large farms tend to be more efficient (Latruffe, et al., 2004; Burja & Burja, 2016). In the past decades, there is a reduction in the number of farms with a tendency to have an increased average farm size (Eastwood, et al., 2010; Piet, et al., 2012). There is an untested belief in Kosovo that farmers having small-sized greenhouse farms may be more likely to receive a grant than farmers with large-sized greenhouse farms. First, it may be correct as MAFRD's objectives over the years have been stable to further promote the greenhouse vegetable production and where most of the farms of this subsector have been relatively small-sized. Second, the provision of new greenhouses to this group of farmers can further support increases in yields and incomes. To quantify this assumption, it may be important to examine the effect of the farmer's greenhouse area in $m^2$ in the matching methods.

Another potentially important covariate is distance to market. For example, farmers' markets bring consumers closer to producers (Ling & Newman, 2011), and the farmer's distance to the market in km may impact the quantity and when the produce is sold. Distance from farm to market may be an important factor determining the farmer's access to the product (output) markets (Ahmed, et al., 2016). Nevertheless, Kosovo is a small country with most farmers facing no great difficulties to access the farmer's or other markets. Whether farmers growing multiple greenhouse vegetable crops or those concentrated in a single vegetable crop per season can be more likely to receive a grant may be among the covariates considered in the matching methods. In Kosovo, farmers tend to grow other crops to ensure that if one crop fails, income may be still generated from the other cultivated crop. There are farms that practice crop diversification because there may be a possibility to achieve family food security (Abdulkadri & Ajibefun, 1998; Ogundari, 2009). In fact, vegetable farms practicing multi-cropping were found to consume less farm inputs such as water, diesel and electricity (Li, et al., 2018). In addition, farmers with five crops have better resource-efficiency than those with less than five crops (Ogundari, 2009). However, reasons may vary as why a farmer grows a single crop or multiple crops.

## Data

### *Summary of Data*

The data for the study were obtained from surveyed greenhouse tomato and pepper farmers in Kosovo from June to August 2017. A sample covering 136 greenhouse farms were surveyed in regions[6] of Prishtina, Ferizaj, Gjilan, Prizren, Gjakova, Peja, and Mitrovica. From the sample of surveyed greenhouse farmers, 94[7] were producing tomatoes and 42 were producing peppers. To ensure an appropriate data collection process, two research surveys[8] were developed to interview greenhouse tomato and pepper farmers. There were three phases to gather the study's data from the field: (a) prioritize municipalities and villages with a greater number of producers growing greenhouse peppers and tomatoes; (b) interviews with greenhouse farmers over the age of 18; and (c) data analysis. Greenhouse vegetable production was chosen because it represents the least theoretically and empirically explored subsector of agriculture and yet with a high economic potential in Kosovo. As greenhouse farming becomes more complex and competition becomes fiercer from imports, it is crucial

[6] During the field visits, this study did not find many commercial greenhouse pepper farmers in the region of Prizren. Thus, Prizren is the only missing region. However, note that in this region there may be pepper production which can be found mainly in open-fields.

[7] Out of the 94 greenhouse farmers producing tomatoes, there were seven of them who were producing simultaneously tomatoes and peppers.

[8] These surveys received approval from the University of Arkansas Institutional Review Board (Approval number 17-04-678) and were carried out with collaboration by faculty at the University of Prishtina "Hasan Prishtina" in Kosovo.

to learn the effects of government grant programs in the vegetable production. One limitation of the data collected was the relatively small sample of the greenhouse pepper farmers.

*Covariate Definitions and Descriptive Statistics*

Age, education in years, yield in kilograms (kg), greenhouse value in euros, greenhouse area in square meters ($m^2$), distance to the market in kilometers (km), and other greenhouse crops grown may be chosen from above to evaluate the greenhouse tomato and pepper seasonal income differences among grantees and non-grantees of the Ministry of Agriculture, Forestry and Rural Development (MAFRD) and the possible participation in a grant program.

In the survey data, the age of a greenhouse farmer was reported in years at the time of the interview. The survey sample reflected a mixture of young and old farmers. In the last 10 years, many new greenhouses have been constructed throughout Kosovo. Some of which were owned by young farmers. Despite a relatively low mean difference in age prior to matching, there were greenhouse tomato farmers as young as 20 years old and as old as 58 years old. Among greenhouse pepper farmers the youngest farmer surveyed was 27 years old and the oldest was 58 years old. Education was reported in years. In this sample, greenhouse pepper farmers appeared to have on average more years of education than the greenhouse tomato farmers. There were greenhouse tomato and pepper farmers with as low as 8 years of education, and as high as 20 years of education.

Often greenhouse farmers' yields reported in kg may dictate the success of a greenhouse farm in Kosovo. A preliminary review of the data showed that there were greenhouse tomato grantees who had slightly more than three times higher yields relative to the non-grantees. Comparatively, greenhouse pepper grantees had two times higher yields than non-grantees. A covariate with stark differences observed in the sample was greenhouse value reported in euros. Greenhouse tomato farmers had a greenhouse mean value of 14,597 euros compared to a mean of 16,729 euros among greenhouse pepper farmers. From the dataset, it was found that a greenhouse tomato grantee had on average a slightly more than four times higher euro value of the greenhouse compared to a non-grantee. In contrast, a greenhouse pepper grantee had on average a greenhouse in euro value of slightly more than one times compared to that of a non-grantee. The mean variations in the greenhouse euro values prior to matching indicate differences between the two groups of farmers. When observing the covariate distance to the market in km, the statistics showed that a greenhouse tomato grantee had on average a nearly two times higher distance to the market compared to a non-grantee. While, a greenhouse pepper grantee had on average a one times lower distance to the market than a non-grantee. However, there were greenhouse tomato farmers who had only a distance of 0.30 km from the market and greenhouse pepper farmers with a minimum distance of 3 km. Greenhouse farmers with a distance to the market of over 65 km were only few and in the remotest areas of Kosovo.

The covariate other greenhouse crops grown was reported as the number of crops cultivated with greenhouse tomatoes and peppers. On average, greenhouse pepper farmers had three crops grown in the same season compared to two crops among greenhouse tomato farmers. Examining the variable greenhouse area in $m^2$, a greenhouse tomato grantee had on average a greenhouse with an area of more than three times of a non-grantee. While, a greenhouse pepper grantee had on average a greenhouse with an area of slightly more than one times of a non-grantee. There were greenhouse tomato small-sized farms with a greenhouse area as low as 100 $m^2$, and large-sized farms as high as 8,500 $m^2$. While, small-sized greenhouse pepper farms had greenhouse areas with a minimum of 200 $m^2$, and large-sized farms with a maximum of 6,000 $m^2$. The outcome of the matching method may be a farmer’s seasonal income. On average, greenhouse pepper farmers made 204 euros more than the greenhouse tomato farmers over the course of a growing season. This study may consider the use of this variable separately for each vegetable crop in the matching methods. Table 3.1 provides descriptive statistics of each covariate and outcome.

Table 3.1. Descriptive Statistics of Covariates and Outcomes

| Statistic | Unit | N | Mean | CV | Min | Median | Max |
|---|---|---|---|---|---|---|---|
| *Tomato covariates (x)* | | | | | | | |
| Age | years | 94 | 40 | 0.22 | 20 | 42 | 58 |
| Education | years | 94 | 11 | 0.30 | 8 | 8 | 20 |
| Yield | kg | 94 | 9,817 | 1.07 | 900 | 5,000 | 50,000 |
| Greenhouse value | euro | 94 | 14,597 | 1.30 | 550 | 7,650 | 92,000 |
| Greenhouse area | $m^2$ | 94 | 899 | 1.28 | 100 | 500 | 8,500 |
| Distance to the market | km | 94 | 18 | 1.03 | 0.30 | 12 | 85 |
| Other crops grown | number | 94 | 2 | 0.36 | 1 | 2 | 4 |
| *Pepper covariates (x)* | | | | | | | |
| Age | years | 42 | 41 | 0.19 | 27 | 40 | 58 |
| Education | years | 42 | 12 | 0.27 | 8 | 12 | 20 |
| Yield | kg | 42 | 4,648 | 0.78 | 650 | 3,550 | 16,000 |
| Greenhouse value | euro | 42 | 16,729 | 1.43 | 1,500 | 10,000 | 150,000 |
| Greenhouse area | $m^2$ | 42 | 898 | 1.00 | 200 | 700 | 6,000 |
| Distance to the market | km | 42 | 25 | 0.84 | 3 | 15 | 84 |
| Other crops grown | number | 42 | 3 | 0.33 | 1 | 3 | 4 |
| *Outcomes (y)* | | | | | | | |
| Tomato seasonal income | euro | 94 | 4,354 | 0.97 | 146 | 3,142 | 20,590 |
| Pepper seasonal income | euro | 42 | 4,558 | 0.77 | 675 | 3,680 | 16,310 |

Note: N, number of observations; CV, coefficient of variation which is defined as follows: $CV = \frac{\sigma}{\mu}$ that is shown as a coefficient in this study with σ indicating the covariate's standard deviation and μ the covariate's mean value. In addition, note that the variation in the greenhouse value is higher among greenhouse tomato farmers. Because the outlier with a value of 150,000 influenced the variation in the greenhouse value among greenhouse pepper farmers to appear larger.

## Results

*Overview*

An exhaustive search algorithm was performed in R programming coupled with an extensive literature review under Kosovo's context to identify an appropriate set of covariates for the matching methods. Employing these techniques, education in years was found as a covariate with no contribution to improve the Akaike information criterion (AIC) of the binary logistic regression models. It was statistically insignificant when it was included in the model for the greenhouse tomatoes (p-value = 0.37), and it was also insignificant in the model for the greenhouse peppers (p-value = 0.48). Therefore, it was not incorporated in the final covariate set. Table 3.2 shows the results and selected covariates with the use of binary logistic regression for both greenhouse tomato and pepper farms. Age, yield in kg, greenhouse value in euros, greenhouse area in square meters ($m^2$), and distance to the market in kilometers (km) were five of the six variables chosen for the use in the matching analysis. The last covariate used in the matching methods was other greenhouse crops grown. This covariate despite the low statistical significance was important to contribute to the overall covariate balance in the matching methods as farmers had a various number of crops grown.

Table 3.2. Logistic Regression Results

| | Greenhouse Tomato Model | | | | Greenhouse Pepper Model | | | |
|---|---|---|---|---|---|---|---|---|
| | | | 95 % CI | | | | 95 % CI | |
| Variable | β (SE) | OR | Lower | Upper | β (SE) | OR | Lower | Upper |
| Age in years | 0.22** (0.09) | 1.248 | 1.083 | 1.548 | 0.22** (0.11) | 1.244 | 1.035 | 1.632 |
| Yield in kg | 0.0003** (0.0001) | 1.000 | 1.000 | 1.001 | 0.001*** (0.0003) | 1.001 | 1.000 | 1.002 |
| Greenhouse value in euros | 0.0004** (0.0002) | 1.000 | 1.000 | 1.001 | 0.002 (0.0001) | 1.000 | 1.000 | 1.001 |
| Greenhouse area in $m^2$ | -0.004** (0.002) | 0.996 | 0.992 | 0.999 | -0.005 (0.004) | 0.995 | 0.985 | 0.999 |
| Distance to the market in km | 0.0623** (0.03) | 1.064 | 1.010 | 1.141 | -0.083* (0.05) | 0.920 | 0.814 | 0.998 |
| Other crops grown | 0.57 (0.99) | 1.770 | 0.245 | 15.273 | 1.80* (1.05) | 6.038 | 1.036 | 80.703 |
| (Constant) | -19.56*** (1.893) | | | | -17.44** (7.22) | | | |
| Observations | 94 | | | | 42 | | | |
| Log Likelihood | -12.65 | | | | -10.07 | | | |
| Akaike Inf. Crit. | 39.29 | | | | 34.14 | | | |
| Nagelkerke $R^2$ | 0.79 | | | | 0.69 | | | |

Note: β, regression coefficient; SE, standard error; OR, odds ratio; CI, confidence interval. The statistical significance of the variables is represented by *p<0.1; **p<0.05; ***p<0.01. Education in years not shown in the model was found statistically insignificant (p-value = 0.37) in the greenhouse tomato model and in the greenhouse pepper model (p-value = 0.48).

After the selection of the covariates, two dimensions of this study explained the differences existing between greenhouse tomato and pepper farmer recipients of the financial support in the form of grants from MAFRD relative to the non-recipients. First, an analysis of any potential covariate balance was provided under the method of propensity score matching (PSM) employing logistic regression as the distance measure. This method demonstrated that matching made worse off the covariate mean differences particularly between greenhouse pepper grantees and non-grantees. In addition, the seasonal income difference emerged as positive for both greenhouse tomato and pepper grantees. Nevertheless, the result revealed no statistical significance. This may be in part due to the new presence of MAFRD grant programs and the discrepancy in the number of grantees relative to the non-grantees. Another reason may be the small sample of farmers for use of PSM with logistic regression as the distance measure. In this line of reasoning, Polson (2017) stated that the propensity score analysis including a small sample size may hinder the possibility to obtain statistically significant results.

Second, genetic matching enabled a comparison of the results before and after matching for a potential disparity between treatment and control groups. This method found a significant improvement of the covariate balance from matching among greenhouse tomato and pepper farmers. Given the nature of the genetic matching that searches in an automatic manner to find a relevant matching of the treatment and control groups, it provided higher statistical significance. A further analysis indicated that the difference in seasonal income among greenhouse tomato grantees and non-grantees was positive and significant at the 5 percent confidence interval. Although no seasonal income difference was found significant among greenhouse pepper treatment and control groups, genetic matching allowed a partial decrease for some of the covariate standardized mean differences (SMDs).

*Genetic Matching Results*

Table 3.3 shows the outcome of the genetic matching method as an adjusted and unadjusted variable for both the greenhouse tomato and pepper farmers. The result indicated a positive and significant average treatment effect on the treated (ATT) of the seasonal income for the greenhouse tomato grantees. A positive and adjusted estimate of 1,777 euros in income per growing season was registered for grantees relative to non-grantees. The higher seasonal income of the grantees may be explained by the financial support received from MAFRD. It was possible to expect that the level of the seasonal income increases with a greenhouse tomato farmer's participation in a MAFRD grant program. The objective of the MAFRD grants was to promote the greenhouse vegetable production, and this study considered that the promotion of this subsector among other impacts may have an impact on farmers' income. However, no statistically significant differences in income estimates were revealed for the greenhouse pepper grantees.

Table 3.3. The Average Treatment Effect on the Treated under Genetic Matching

| | | Greenhouse tomato grantees | | | Greenhouse pepper grantees | | |
|---|---|---|---|---|---|---|---|
| Outcome variable | Unit | Mean | AI SE | p-value | Mean | AI SE | p-value |
| *Seasonal income* | | | | | | | |
| Adjusted estimate | euro | 1,777.1 | 867 | 0.04** | 1,704.2 | 1,862 | 0.36 |
| Unadjusted estimate | euro | 1,777.1 | 613 | 0.003*** | 1,704.2 | 1,428 | 0.23 |

Note: AI SE, Abadie-Imbens standard error. The estimation procedure included the average treatment effect on the treated (ATT). The statistical significance of the variables is represented by *p<0.1; **p<0.05; ***p<0.01.

Figure 3.1 and Table 3.4 suggested under genetic matching for the greenhouse tomato grantees relative to the non-grantees that the SMD of the covariate yield in kg decreased from 1.38 to 0.29 noting that matching may minimize the mean differences. Farmers with lower yields often had smaller greenhouse areas. The difference in yields for a non-grantee farmer with a 500 $m^2$ greenhouse was nearly twice less than that of a grantee's 1000 $m^2$ greenhouse area. The reduction of 79 percent registered in the SMD of the covariate yield in kg was important given the high initial discrepancy between grantees and non-grantees. There were greenhouse tomato grantees with reported yields of slightly more than three times of the non-grantees' yields implying a difficulty to achieve an SMD below the recommended threshold of 0.1. Although the matching result of the covariate showed an improvement, yet the greenhouse tomato grantees and the non-grantees remained with a significant divergence in yields. The SMD of the covariate greenhouse area with a decrease from 1.04 to 0.04 and a 96 percent reduction in its SMD appeared to be the most relevant matching result. The result may suggest that considering this covariate, a grantee from MAFRD may be properly matched with a non-grantee in the production of greenhouse tomatoes.

In contrast, the SMD of the covariate distance to the market in km decreased from 0.41 to 0.35 with a reduction of only 15 percent. This may not indicate that a grantee can be properly matched with a non-grantee in terms of their distance to the market in km. The inability of the farmers to be matched in terms of the distance to the market in km may be due to improved rural infrastructure. An average distance of 9 km may not necessarily make both groups of farmers different when considering their access to the market. Moreover, the SMD of the covariate other greenhouse crops grown also decreased from 0.68 to 0.31 but with a significant reduction of 54 percent in its SMD. Farmers tend to grow multiple crops with the objective to increase crop diversity. One reason may be that if one of the crops failed during the production process, farmers may rely on the second crop as a source of income. However, given that the covariate was not below the recommended threshold, a grantee may not be matched with a non-grantee in terms of the number of other crops grown over the course of a season. In contrast, the covariate age increased without a significant impact on the matching quality. This implied that age prior and after matching was insignificant and with no discrepancy in the mean values. Therefore, age per se may not dictate a greenhouse tomato farmer's participation in a MAFRD grant program.

Referring to Figure 3.2 and Table 3.4, the greenhouse pepper grantees relative to the non-grantees reflected a decrease in the SMD of the covariate yield from 1.07 to 0.53. The 50 percent reduction in its SMD may favor a match between the greenhouse grantees and the non-grantees. This noted that after matching, the difference in yields had no statistically different mean values between the subjects. Nevertheless, still more improvement was required to achieve an SMD below the recommended threshold of 0.1. In comparison to the greenhouse tomato subjects, on average, greenhouse pepper subjects had a lower variation in mean values only for the covariate yield in kg. Despite the comparatively lower variation in the mean values, this covariate may have favored a match as grantees and non-grantees reported a gap in yields.

The remaining variables showed a mixture of increases and decreases in SMDs, however, with no proper matching. Increases in SMDs were registered for the covariates age, distance to the market in km, greenhouse area in $m^2$, and greenhouse value in euros. These increases in SMDs revealed that matching provided no approximation in covariate mean values. Although a decrease in the SMD of the covariate other greenhouse crops grown was registered, it was still marginal.

Table 3.4. Covariate Balance Results using Genetic Matching

| Variable | Before matching with a ratio 3:1 | | | | After matching with a ratio 3:1 | | | |
|---|---|---|---|---|---|---|---|---|
| | Grantees | Non-grantees | p-value | d | Grantees | Non-grantees | p-value | d |
| *Tomatoes* | N = 94 | | | | N = 16 Unweighted N = 16 | | | |
| Age | 39.68 | 39.84 | 0.95 | 0.02 | 39.68 | 40.00 | 0.93 | 0.03 |
| Yield | 23,213 | 7,069 | <0.001 | 1.38 | 23,213 | 19,050 | 0.002 | 0.29 |
| Greenhouse value | 42,720 | 8,828 | <0.001 | 1.54 | 42,720 | 22,956 | 0.004 | 0.82 |
| Greenhouse area | 2,238 | 624 | 0.009 | 1.04 | 2,238 | 2,172 | 0.89 | 0.04 |
| Distance to the market | 25.21 | 16.25 | 0.21 | 0.41 | 25.21 | 36.31 | 0.31 | 0.35 |
| Other crops grown | 2.50 | 2.03 | 0.01 | 0.68 | 2.50 | 2.19 | 0.25 | 0.31 |

| Variable | Before matching with a ratio 3:1 | | | | After matching with a ratio 3:1 | | | |
|---|---|---|---|---|---|---|---|---|
| | Grantees | Non-grantees | p-value | d | Grantees | Non-grantees | p-value | d |
| *Peppers* | N = 42 | | | | N = 10 Unweighted N = 10 | | | |
| Age | 41.70 | 40.78 | 0.77 | 0.11 | 41.70 | 35.50 | 0.03 | 0.76 |
| Yield | 7,910 | 3,629 | 0.03 | 1.07 | 7,910 | 5,748 | 0.10 | 0.53 |
| Greenhouse value | 25,190 | 14,084 | 0.13 | 0.51 | 25,190 | 17,050 | 0.13 | 0.64 |
| Greenhouse area | 1,210 | 800 | 0.08 | 0.53 | 1,210 | 950 | 0.07 | 0.73 |
| Distance to the market | 24.10 | 25.22 | 0.88 | 0.05 | 24.10 | 29.7 | 0.52 | 0.29 |
| Other crops grown | 2.80 | 2.44 | 0.23 | 0.44 | 2.80 | 2.70 | 0.32 | 0.14 |

Note: P value, T-test p value; N, number of observations; Unweighted, indicates several grantees were matched to several non-grantees; d, standardized mean difference is defined as $d = \frac{(\overline{X}_{grantees} - \overline{X}_{non\text{-}grantees})}{\sqrt{\frac{s^2_{grantees} + s^2_{non\text{-}grantees}}{2}}}$, where $\overline{X}_{grantees}$ is the mean value of the grantees and $\overline{X}_{non\text{-}grantees}$ the mean value of the non-grantees. Variance is indicated by $S^2$. The number of 1000 bootstraps was included for the matching balance of the covariates which may provide pertinent Kolmogorov-Smirnov test values as recommended by Sekhon (2011). Statistics shown in the table are mean values for the grantees and control group non-grantees.

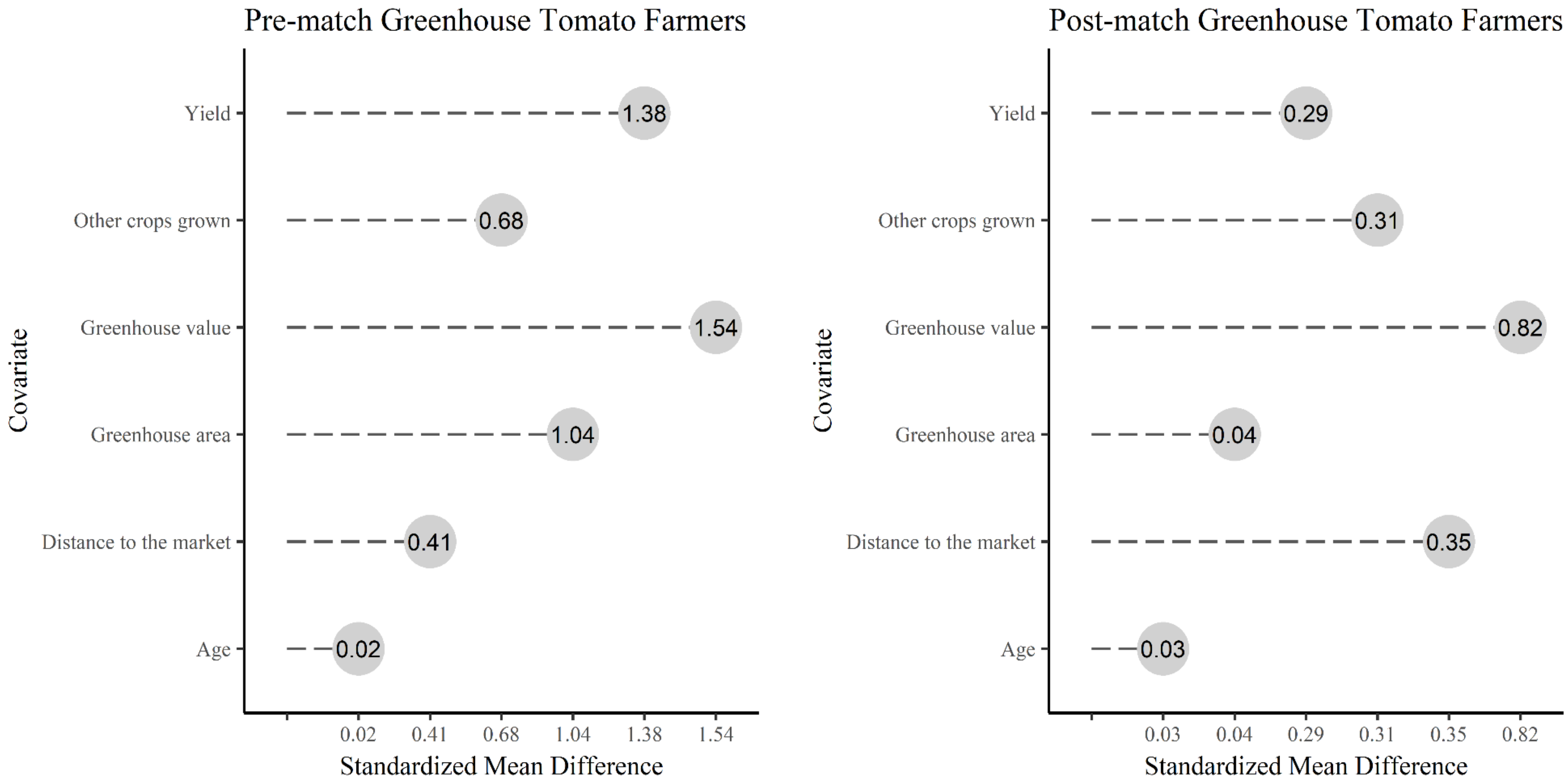


Figure 3.1. Pre- and Post-match Greenhouse Tomato Standardized Mean Differences (SMDs) under Genetic Matching.
Note: Each circle shows the covariate's SMD. The increase or decrease in a SMD is found through comparing a covariate's SMD from the pre-match to the after-match phase.

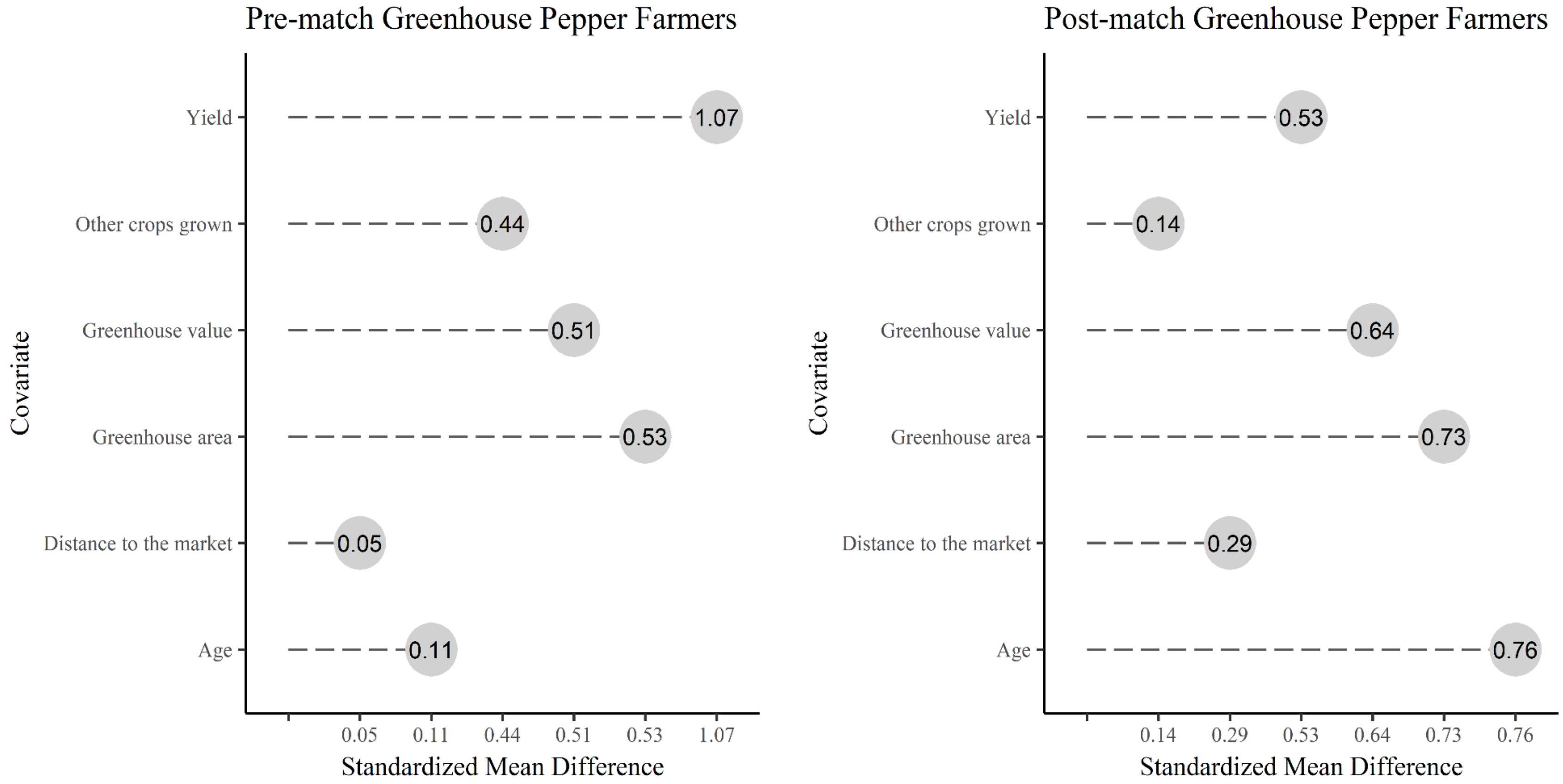


Figure 3.2. Pre- and Post-match Greenhouse Pepper Standardized Mean Differences (SMDs) under Genetic Matching.
Note: Each circle shows the covariate's SMD. The increase or decrease in a SMD is found through comparing a covariate's SMD from the pre-match to the after-match phase.

*Propensity Score Matching Results with Logistic Regression*

With the use of PSM together with logistic regression as the distance measure and the ATT, this study found less significant covariate balance relative to the genetic matching method employing the same evaluation criteria. The criteria included the estimation of ATT with replacement and a matching ratio of 3:1. After the matching procedure, the overall assessment of the covariate balance indicated a decrease in the SMDs between the treatment and the control groups of farmers. This was among the objectives of the study, as lower SMDs may provide a better covariate balance between the treatment and control groups of farmers. However, not all the covariates progressed towards balance, and the SMDs were not all within the preferred 0.1 threshold.

It is notable that there were farmers who received grants to construct new greenhouses, which had in some cases two to three times higher euro values compared to the traditional greenhouses used by the grant non-recipients. During the model specification in R programming it was difficult to use the right matching criteria that address fully the large mean differences of some of the covariates. To address this difficulty, this study was led to the understanding that the genetic matching method may be a relevant strategy to search automatically for proper matching. However, a comparison of the results with PSM using logistic regression as the distance measure was important to examine the quality of matching. The results showed that the control and treatment means before matching for the covariates yield in kg, greenhouse value in euros, and greenhouse area in $m^2$ had more statistically significant differences compared to the age, distance to the market in km, and other greenhouse crops grown. There were farmers who reported relatively high values for a given covariate and farmers who reported comparatively low values. When analyzing the covariates, this led to a significant gap in the mean values which in part affected the process of not having all the covariates with SMDs in the region of less than 0.1. The control and treatment groups of the greenhouse tomato subjects had an overall less discrepancy in mean values after matching compared to the control and treatment groups of the greenhouse pepper subjects. An overview of the distribution of the propensity scores is shown in Figure 3.3 through referring to the logistic regression model presented in Table 3.2. It suggested that greenhouse tomato farmers before matching had a positive skew with a concentration of the propensity scores below the mean. This was coupled with a density curve towards lower propensity score levels. It is indicated that greenhouse pepper farms show relatively low probabilities for a potential participation in the MAFRD grant programs.

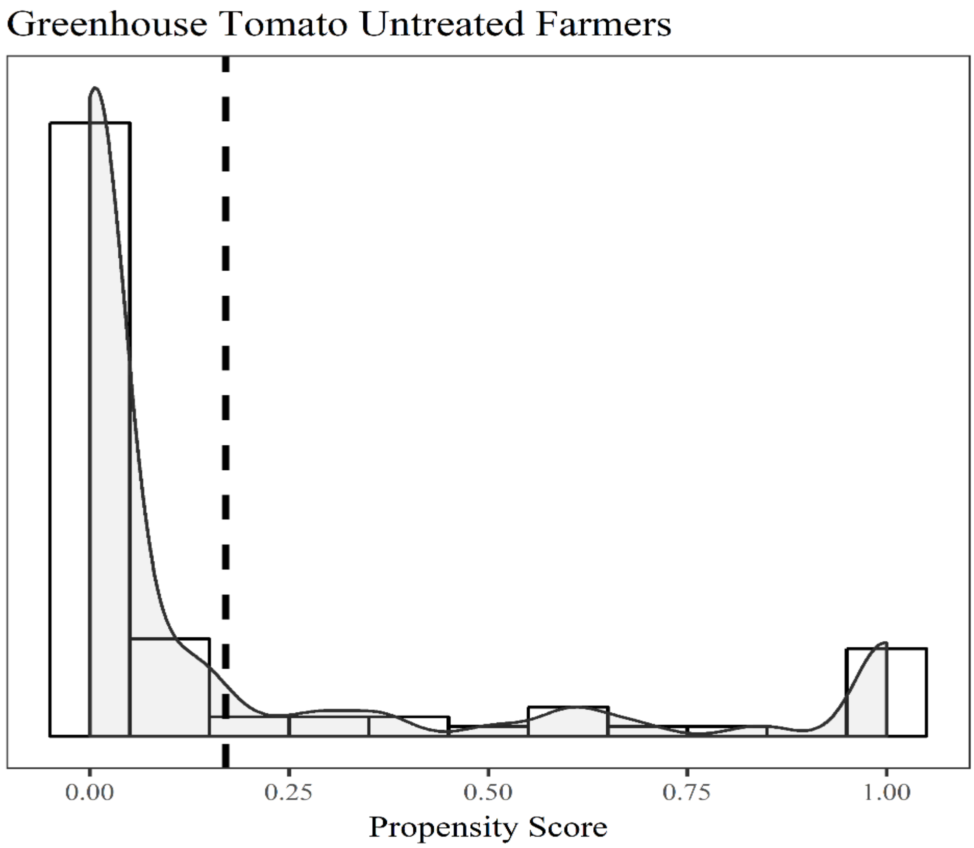


Figure 3.3. Greenhouse Tomato Untreated Farmers under PSM with Logistic Regression.
Note: The vertical and dashed line represents the mean line, while the shaded area shows the density over the histogram. Low propensity scores before matching are registered for a participation in a grant program.

The distribution of the propensity scores are shown in Figure 3.4 by using the logistic regression model presented above in Table 3.2. A relatively positive skew is registered among the covariates. This applied to the greenhouse pepper farmers before matching. A concentration of the probabilities representing propensity scores was registered below the mean. The density curve for the greenhouse pepper farmers is slightly less positively skewed compared to the greenhouse tomato farmers' density curve. However, it demonstrates nearly the same probabilities for a potential participation in the MAFRD grant programs.

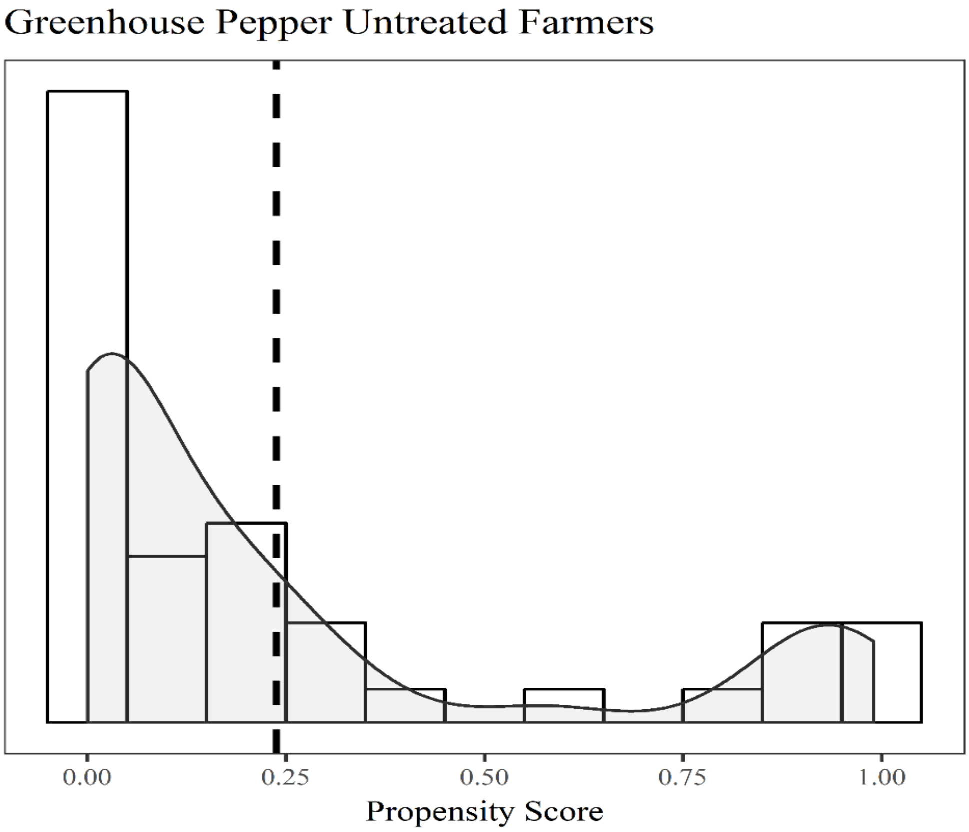


Figure 3.4. Greenhouse Pepper Untreated Farmers under PSM with Logistic Regression.
Note: The vertical and dashed line represents the mean line, while the shaded area shows the density over the histogram. Relatively low propensity scores before matching are registered for a participation in a grant program.

Overall, both logistic regression models presented sufficient statistical significance to be included in the PSM method with logistic regression as the distance measure. However, the small sample size of 42 greenhouse pepper farmers may be a limitation leading to higher SMDs among the covariates after matching. While evaluating the effects of the greenhouse tomato grantees relative to the non-grantees, there was no statistical significance on the farmer's seasonal income. With the use of PSM with logistic regression as the distance measure, greenhouse tomato farmers' seasonal income with an adjusted estimate of 775.71 euros was insignificant. Likewise, seasonal income with an adjusted estimate of 3,655 euros was insignificant for the greenhouse pepper farmers. Although insignificant, the impact of grants on greenhouse pepper farmers' seasonal income emerged to be positive by two of the implemented matching techniques. Table 3.5 shows the outcome as an adjusted and unadjusted estimate in euro values for both the greenhouse tomato and pepper farmers. With the PSM method using logistic regression as the distance measure, it was noted that the unadjusted estimate of 3,655 euros is statistically significant at the 5 percent confidence interval for the greenhouse pepper farmers. A potential implication of this may be that with the increase of the sample size, the adjusted estimate of the seasonal income may become statistically significant at the 5 percent confidence interval.

Table 3.5. The Average Treatment Effect on the Treated under PSM with Logistic Regression

| | | Greenhouse tomato grantees | | | Greenhouse pepper grantees | | |
|---|---|---|---|---|---|---|---|
| Outcome variable | Unit | Mean | AI SE | p-value | Mean | AI SE | p-value |
| *Seasonal income* | | | | | | | |
| Adjusted estimate | euro | 775.71 | 2,634 | 0.77 | 3,655 | 2,323 | 0.12 |
| Unadjusted estimate | euro | 775.71 | 1,694 | 0.64 | 3,655 | 1,637 | 0.03** |

Note: AI SE, Abadie-Imbens standard error. The estimation procedure included the average treatment effect on the treated (ATT) under the method of PSM with logistic regression. The statistical significance of the variables is represented by *p<0.1; **p<0.05; ***p<0.01.

The covariate balance becomes less stable compared to genetic matching with covariates having higher than 0.1 SMDs. The farmer's age, distance to the market in km, and other greenhouse crops grown were three covariates that did not contribute to the model as they showed increases in the distance of their mean values after performing matching. The mean comparisons between the greenhouse tomato and pepper grantees relative to the non-grantees before matching presented stark differences. Nevertheless, the after matching t-test p value results indicated that not all the covariates were significant. Table 3.6 and Figure 3.5 suggested that for the greenhouse tomato grantees relative to the non-grantees, the SMD of the covariate yield in kg decreased from 1.38 to 0.14 with a 90 percent reduction. An implication of this was that compared to genetic matching, the covariate yield in kg had a higher matching quality. Prior to matching, the greenhouse pepper grantees and non-grantees had statistically significant mean differences in yields. After matching, the same subjects had insignificant differences. This may suggest that through using both matching techniques examined in this study, a matching may be possible between a grantee with high yields and a non-grantee with relatively lower yields.

In the same manner, the SMD of the covariate greenhouse value decreased from 1.54 to 0.90 with a 42 percent reduction. Before matching, this covariate had marked differences in the mean values which after matching became insignificant. This showed that using this method farmers with greenhouses that have improved designs and structures may be matched with farmers that have small-sized and less costly greenhouses. Among some of the effects produced by the MAFRD grant programs was in fact a major gap in the greenhouse euro value among the greenhouse tomato grantees and non-grantees. In this line of reasoning, the ability to provide a match between the two groups of farmers may be of importance and possibly achieved. The SMD of the covariate other greenhouse crops grown decreased from 0.68 to 0.56 with a marginal reduction of 18 percent. Prior to the matching procedure, it had a significant difference in the mean value which after matching was insignificant. The remaining covariates registered a mixture of increases and decreases in SMDs, however, with no significant differences. An increase in the SMD was found for the covariates age, while decreases in SMDs were found for the covariates greenhouse area and distance to the market in km. Despite the changes in SMDs, they provided no significance.

For the greenhouse pepper treatment and control groups, Table 3.6 indicated that all the covariate SMDs grew significantly except the covariate other greenhouse crops grown. Still, the covariate other greenhouse crops grown showed no statistical significance in the mean values prior to matching. This result is noted in Figure 3.6 where the greenhouse pepper grantees relative to the non-grantees overall were made worse off from matching. In fact, they showed no statistical differences in the covariate mean values before matching except the covariate yield in kg, which marked, however, an increase in its SMD.

Table 3.6. Covariate Balance Results using PSM with Logistic Regression

| Variable | Before matching with a ratio of 3:1 | | | | After matching with a ratio of 3:1 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Grantees | Non-grantees | p-value | d | Grantees | Non-grantees | p-value | d | OR |
| *Tomatoes* | N = 94 | | | | N = 16<br>Unweighted N = 48 | | | | |
| Age | 39.69 | 39.85 | 0.95 | 0.02 | 39.69 | 46.50 | 0.10 | 0.65 | 1.25** |
| Yield | 23,213 | 7,069 | <0.001 | 1.38 | 23,213 | 21,392 | 0.69 | 0.14 | 1.00** |
| Greenhouse value | 42,720 | 8,828 | <0.001 | 1.54 | 42,720 | 21,021 | 0.02 | 0.90 | 1.00** |
| Greenhouse area | 2,238 | 624 | 0.009 | 1.04 | 2,238 | 2,166 | 0.90 | 0.04 | 0.99** |
| Distance to the market | 25.21 | 16.25 | 0.21 | 0.41 | 25.21 | 36.49 | 0.31 | 0.37 | 1.10** |
| Other crops grown | 2.50 | 2.03 | 0.01 | 0.68 | 2.50 | 1.92 | 0.13 | 0.56 | 1.77 |
| Variable | Before matching with a ratio of 3:1 | | | | After matching with a ratio of 3:1 | | | | |
| | Grantees | Non-grantees | p-value | d | Grantees | Non-grantees | p-value | d | OR |
| *Peppers* | N = 42 | | | | N = 10<br>Unweighted N = 30 | | | | |
| Age | 41.7 | 40.78 | 0.77 | 0.11 | 41.7 | 48.73 | 0.05 | 1.09 | 1.24** |
| Yield | 7,910 | 3,629 | 0.03 | 1.07 | 7,910 | 3,420 | 0.03 | 1.22 | 1.00*** |
| Greenhouse value | 25,190 | 14,084 | 0.13 | 0.51 | 25,190 | 8,533 | 0.01 | 1.41 | 1.00 |
| Greenhouse area | 1,210 | 800 | 0.08 | 0.53 | 1,210 | 683 | 0.01 | 1.42 | 0.99 |
| Distance to the market | 24.1 | 25.22 | 0.88 | 0.05 | 24.1 | 11.7 | 0.09 | 0.85 | 0.92* |
| Other crops grown | 2.80 | 2.44 | 0.23 | 0.45 | 2.80 | 3.13 | 0.37 | 0.43 | 6.03* |

Note: p-value, T-test p value; N, number of observations; Unweighted, indicates several grantees were matched to several non-grantees; OR, odds ratio; d, standardized mean difference is defined as $d = \frac{(\overline{X}_{grantees} - \overline{X}_{non\text{-}grantees})}{\sqrt{\frac{s^2_{grantees} + s^2_{non\text{-}grantees}}{2}}}$, where $\overline{X}_{grantees}$ is the mean value of the grantees and $\overline{X}_{non\text{-}grantees}$ the mean value of the non-grantees. Variance is indicated by $S^2$. The number of 2000 bootstraps was included for the PSM with logistic regression balance of the covariates only for the greenhouse peppers given its small sample size which may provide pertinent Kolmogorov-Smirnov test values. A minimum of 1000 bootstraps was applied to other models as recommended by Sekhon (2011). Statistics shown in the table are mean values for the grantees and non-grantees. The statistical significance of the variables is represented by *p<0.1; **p<0.05; ***p<0.01.

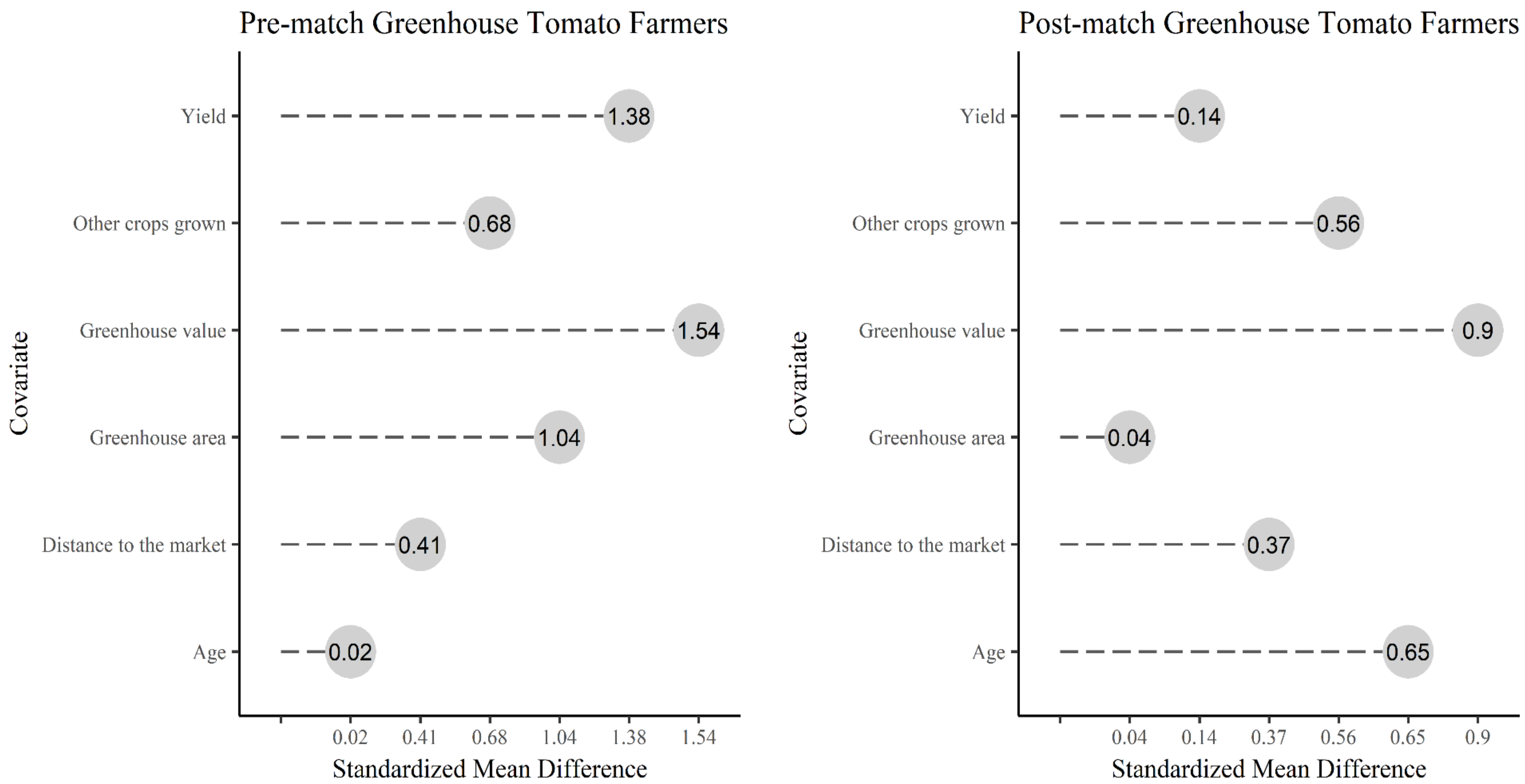


Figure 3.5. Pre- and Post-match Greenhouse Tomato SMDs under PSM with Logistic Regression.
Note: Each circle shows the covariate's SMD. The increase or decrease in a SMD is found through comparing a covariate's SMD from the pre-match to the after-match phase.

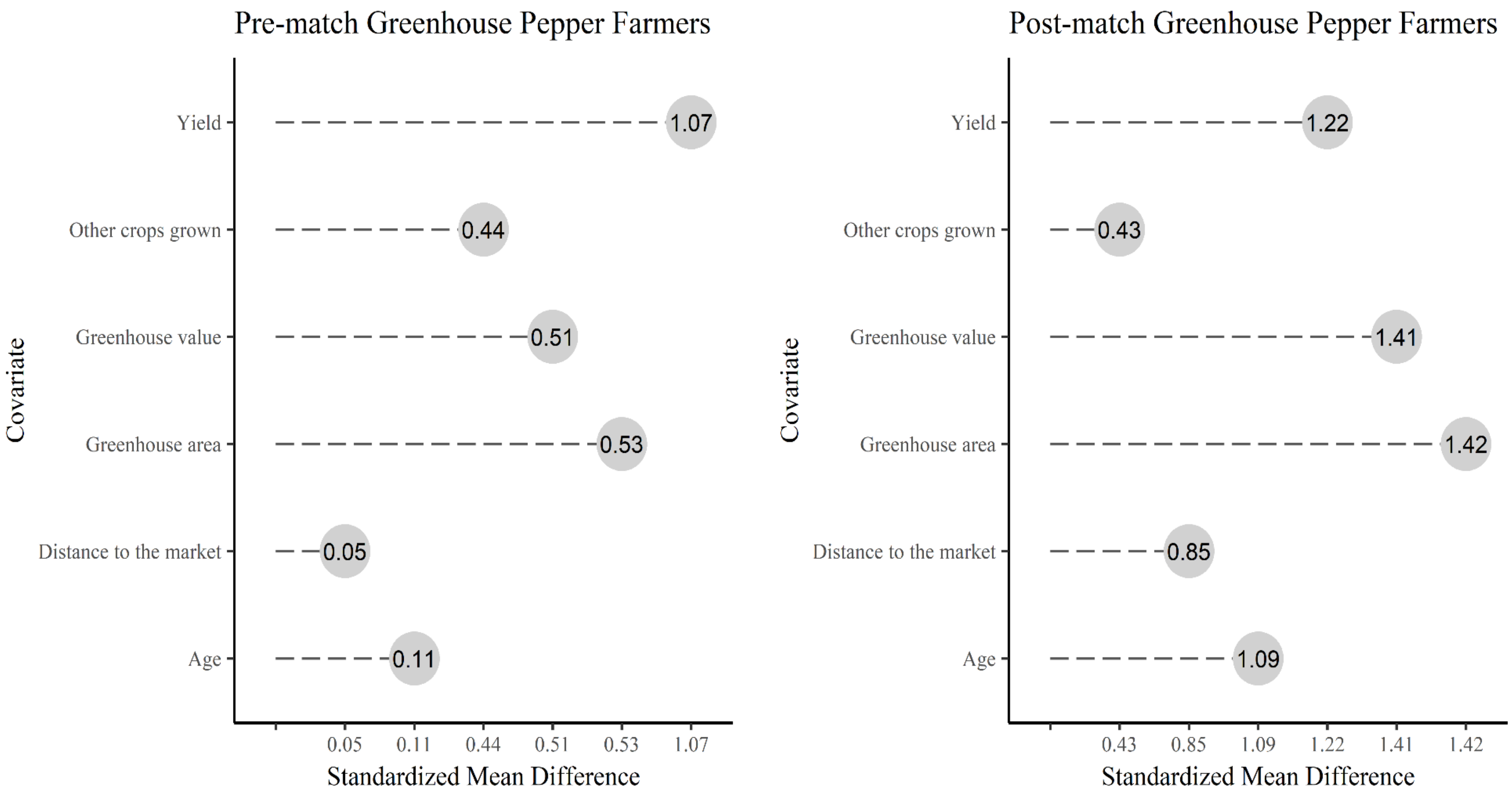


Figure 3.6. Pre- and Post-match Greenhouse Pepper SMDs under PSM with Logistic Regression.
Note: Each circle shows the covariate's SMD. As in the case of the greenhouse tomato farmers, the increase or decrease in a SMD is found through comparing a covariate's SMD from the pre-match to the after-match phase.

## Conclusion

The presence of the government grant programs as an agricultural policy may provide the possibility to promote Kosovo's greenhouse production. Each year more and more farmers apply to the grant program titled "Measure 101: Investments in physical assets and agricultural households" (among other funding initiatives) provided by the Ministry of Agriculture, Forestry and Rural Development (MAFRD) in Kosovo. This is an investment program that funded successful applications for new greenhouses, support for open-field production and/or storage warehouses for the vegetable crops (MAFRD, 2016).

This study's findings support the notion that farmers' participation in this grant program may be related to the farm's yield in kilogram (kg), greenhouse value in euros, and greenhouse area in square meters ($m^2$). Policy researchers in Kosovo may also take interest in the evidence of the positive seasonal income difference of 1,777 euros for the greenhouse tomato grantees relative to the non-grantees. This evidence helps to identify which group of greenhouse farmers are likely to be influenced from the MAFRD grant programs. This study's findings should also be of interest to nonprofit organizations and agencies for development that invest to help MAFRD's efforts in Kosovo for the provision of new and upgraded farm facilities and greenhouses.

Regarding the impact estimates, this study found the genetic matching method with a better overall convergence of the results with our sample of surveyed farmers compared to the propensity score matching (PSM) method using logistic regression as the distance measure. Despite the greenhouse tomato grantees relative to the non-grantees large mean differences prior to matching, the genetic matching method provided a significant improvement in the covariate balance. This demonstrated genetic matching's efficiency of implementation. However, the matching techniques together indicated less significant covariate balance for the greenhouse pepper grantees relative to the non-grantees.

It is notable to acknowledge some limitations of the study. First, although this study identified improvements in the mean values for the control and treatment groups after matching, yet the small sample size of the greenhouse pepper farmers hindered the possibility of achieving covariate SMDs below the preferred 0.1 threshold. Prior to matching, there were significant differences in the covariates of greenhouse area in $m^2$, yield in kg and greenhouse value in euros among greenhouse tomato and pepper farmer groups. The high divergence in mean values for these covariates may have affected the matching quality of the subjects. Second, this study found no evidence of a significant difference in seasonal income among the greenhouse pepper grantees compared to the non-grantees.

In conclusion, these overall results suggest that greenhouse tomato farmer recipients of the grants from MAFRD attained higher incomes per growing season relative to the non-recipients. While, the greenhouse pepper untreated farmers from MAFRD were better off in the pre-matching phase relative to the post-matching phase and with no evidence of a statistically significant difference in the seasonal income.

## References


Abdulkadri, A. O., and I. A. Ajibefun. 1998. "Developing Alternative Farm Plans for Cropping System Decision Making." *Agricultural Systems* 431-442.

Ahmed, U. I., L. Ying, M. K. Bashir, M. Abid, E. Elahi, and M. A. Iqbal. 2016. "Access to Output Market by Small Farmers: The Case of Punjab, Pakistan." *Journal of Animal and Plant Sciences* 787-793.

Akimowicz, M., M. B. Magrini, A. Ridier, J. E. Bergez, and D. Requier-Desjardins. 2013. "What Influences Farm Size Growth? An Illustration in Southwestern France." *Applied Economic Perspectives and Policy* 242-269.

Ali, A., M. Sharif, K. Mahmood, and N. Akmal. 2013. "Determinants of Cherry Production and Marketing in Pakistan: A Propensity Score Matching Approach." *Agricultural Economics Review*.

Arcotrass Gmbh; Vakakis International SA; Eurocare Gmbh; AKI. 2006. *Kosovo Report.* Arcotrass-Consortium.

Birkbak, J., E. A. Stuart, B. D. Lind, P. Qin, E. Stenager, K. J. Larsen, . . . A. Erlangsen. 2016. "Psychosocial Therapy and Causes of Death after Deliberate Self-harm: A Register-based, Nationwide Multicentre Study Using Propensity Score Matching." *Psychological Medicine* 3419-3427.

Burja, C., and V. Burja. 2016. "Farms Size and Efficiency of the Production Factors in Romanian Agriculture." *Ekonomika Poljoprivrede* 361-374.

Caliendo, M., and S. Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 31-72.

Cochran, W. G., and D. B. Rubin. 1973. "Controlling Bias in Observational Studies: A Review." *Sankhyā: The Indian Journal of Statistics* 417-446.

D'Agostino, R. B. 1998. "Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-randomized Control Group." *Statistics in Medicine* 2265-2281.

Dehejia, R. H., and S. Wahba. 2002. "Propensity Score Matching Methods for Non-experimental Causal Studies." *Review of Economics and Statistics* 151–161.

Diamond, A., and J. S. Sekhon. 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* 932-945.

Eastwood, R., M. Lipton, A. Newell, R. E. Evenson, and P. L. Pingali. 2010. Farm Size. Handbook of Agricultural Economics.

Fîntîneru, G., and O. Madsen. 2012. "Farmers' Education and Farm Productivity. Evidence from Denmark and from Romania." *AgroLife Scientific Journal* 175-180

Ho, D. E., K. Imai, G. King, and E. A. Stuart, E. A. 2007. “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference.” *Political Analysis* 199-236.

Howarter, S. 2015. “The Efficacy of Propensity Score Matching in Bias Reduction with Limited Sample Sizes.”

International Monetary Fund. 2011. *Staff Report for the 2011 Article IV Consultation and the Initiation of a Staff-Monitored Program.*

Jacovidis, J. N. 2017. “Evaluating the Performance of Propensity Score Matching Methods: A Simulation Study.” PhD Dissertation, James Madison University

Jaggard, K. W., A. Qi, and E. S. Ober. 2010. “Possible Changes to Arable Crop Yields by 2050.” *Philosophical Transactions: Biological Sciences* 2835-2851.

Joffe, M. M., and P. R. Rosenbaum. 1999. “Invited Commentary: Propensity Scores.” *American Journal of Epidemiology* 327-333.

Kaciu, S. 2008. “Study of the current situation of pepper production in Kosovo.”

Kaciu, S., I. Babaj, S. Aliu, and I. Demaj. 2016. “Potential of Protected Vegetable Production in Kosovo and Future Perspectives.” *Acta Hortic* 461-466.

Kerolli-Mustafa, M., and E. Gjokaj. 2016. *Kosovo*: Agricultural Policy Development and Assessment.* Joint Research Centre: Institute for Prospective Technological Studies (IPTS).

Kwon, J., K. Kang, G. Kweon, Y. Choi, N. Kang, J. Lee, and H. Rhee. 2006. “Effect of Automatic Ventilation of Greenhouse during Daytime on the Growth and Wilting Occurrence in Watermelon.” *Korean Journal of Horticultural Science and Technology* 138-142.

Latruffe, L., K. Balcombe, S. Davidova, and K. Zawalinska. 2004. “Determinants of Technical Efficiency of Crop and Livestock Farms in Poland.” *Applied Economics* 1255-1263.

Li, L., W. Wu, P. Giller, J. O’Halloran, L. Liang, P. Peng, and G. Zhao. 2018. “Open Accesslife Cycle Assessment of a Highly Diverse Vegetable Multi-cropping System in Fengqiu County, China.” *Sustainability* 10(4).

Ling, C., and L. L. Newman. 2011. “Untangling the Food Web: Farm-to-market Distances in British Columbia, Canada.” *Local Environment* 807-822.

Luellen, J. K., W. R. Shadish, and M. H. Clark. 2005. “Propensity Scores: An Introduction and Experimental Test.” *Evaluation Review* 530-58.

Mahmudul, H. A. 2016. “Farmer’s Education Impact on Farmer’s Wealth in Bangladesh.” *International Journal of Information, Business and Management* 89-92.

Mebane, W. R., and J. S. Sekhon. 1998. “Genetic Optimization Using Derivatives (GENOUD).” *Political Analysis* 187-210.

Miftari, I., B. Hoxha, and E. Gjokaj. 2016. *Kosovo*: Agricultural Policy Brief (CAPB).* Joint Research Centre: Institute for Prospective Technological Studies (IPTS).

Ministry of Agriculture, Forestry and Rural Development. 2016. *Green Report.* Prishtina: Ministry of Agriculture, Forestry and Rural Development.

Ogundari, K. 2009. "Technical Efficiency of Farmers under Different Multi-cropping Systems in Nigeria." *Tropical and Subtropical Agroecosystems* 117-120.

Panda, S. 2015. "Farmer Education and Household Agricultural Income in Rural India." *International Journal of Social Economics* 514-529.

Piet, L., L. Latruffe, C. Le Mouël, and Y. Desjeux. 2012. "How Do Agricultural Policies Influence Farm Size Inequality? The Example of France." *European Review of Agricultural Economics* 5-28.

Polson, K. S. 2017. "Analyzing Electricity Use of Low Income Weatherization Program Participants Using Propensity Score Analysis and a Hierarchical Linear Growth Model." PhD Dissertation, University of Denver.

Ray, D. K., N. Ramankutty, N. D. Mueller, P. C. West, and J. A. Foley. 2012. "Recent Patterns of Crop Yield Growth and Stagnation." *Nature Communications* 1293.

Rosenbaum, P. R., and D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 41-55.

Rubin, D. B. 1977. "Assignment to a Treatment Group on the Basis of a Covariate." *Journal of Educational Statistics* 1-26.

Ruiz, D., D. Stout, and C. Herlihy. 2017. "Use of Genetic Matching in Program Evaluation: The Case of RAD." *Cityscape* 337-349.

Sekhon, J. S. 2011. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R." *Journal of Statistical Software*.

Shapiro, D., R. D. Bollman, and P. Ehrensaft. 1987. "Farm Size and Growth in Canada." *American Journal of Agricultural Economics* 477-483.

Stuart, E. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science : A Review Journal of the Institute of Mathematical Statistics* 1–21.

Stuart, E., and D. B. Rubin. 2008. *Best Practices in Quasi–Experimental Designs: Matching Methods for Causal Inference.* Thousand Oaks, CA: SAGE Publications Ltd.

Tabak, Y. P., M. D. Zilberberg, R. S. Johannes, X. Sun, and L. C. McDonald. 2013. "Attributable Burden of Hospital-onset Clostridium Difficile Infection: A Propensity Score Matching Study." *Infection Control and Hospital Epidemiology* 588-96.

Terza, J. V., A. Basu, and P.J. Rathouz. 2008. "Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics* 531-543.

Wang, Y., L. Xin, X. Li, and J. Yan. 2016. "Impact of Land Use Rights Transfer on Household Labor Productivity: A Study Applying Propensity Score Matching in Chongqing, China." *Sustainability* 1-18.

Yang, S.-f., and D. S. Simbeye. 2013. "Computerized Greenhouse Environmental Monitoring and Control System Based on LabWindows/CVI." *Journal of Computers* 399-408.

Zhong, Z. 2004. "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence." *The Review of Economics and Statistics* 91-107.

## Chapter 4. Conclusion

The objectives of this thesis were to better understand the current production efficiency and ways in which the greenhouse tomato and pepper farmers have been influenced by the government agricultural policies in Kosovo.

The first study aimed to analyze greenhouse tomato and pepper input efficiency use in Kosovo at the farm and regional levels and to determine the external factors that affect efficiency. Data envelopment analysis (DEA) and regression analyses were techniques employed in this study. The regional analysis using the Banker-Charnes-Cooper (BCC) model, found that the region of Prizren is the most efficient in the production of greenhouse tomatoes, and region of Prishtina is the most efficient in the production of greenhouse peppers, respectively. However, improvements in efficiency can be made in both types of production in all regions. These results may be relevant because greenhouse vegetable production is an emerging sector in Kosovo with no proven production practices to increase yields. To achieve technical efficiency (TE) in greenhouse tomato production, regional improvements in input use efficiency are needed by 13 (Prizren) to 41 (Ferizaj) percent. Greenhouse pepper producing regions have more efficient producers. However, in all regions, efficiency can still be enhanced; Prishtina could improve its use of inputs by only 1 percent, while Gjilan could improve by 16 percent.

At a farm level, the BCC model shows that 33 percent of greenhouse tomato farms are fully efficient and only 16 percent under the CCR model. While only 52 percent of the greenhouse pepper farms are fully efficient under the BCC model and 26 percent under the CCR model, respectively. The results also suggest a policy is of vital interest to address the issue of selling greenhouse tomatoes with a price that may jeopardize the financial health of the farms. Region of Prizren with the most concentration of greenhouse tomato farms and region of Gjakova are found to be particularly influenced by the price received per kilogram (kg) of tomatoes from the vegetable wholesalers. Meanwhile, no farmer's complaint is found significant in relation to the price received by greenhouse pepper farms per kg of peppers. This shows that greenhouse pepper farmers on average are more satisfied with the prices received per kg of peppers relative to the greenhouse tomato farmers. Overall, the first study explores the inefficient input use in the production of greenhouse tomatoes and peppers, which is caused by two relevant factors. One factor is the disproportionate use of inputs. The other factor is concerning the unfavorable market conditions particularly from imports and low prices set from vegetable wholesalers. This study revealed that under the given greenhouse tomato and pepper production levels, there is a large opportunity for the technically inefficient farms and regions to improve their use of inputs.

The main contributions of the second study are to expand our understanding whether greenhouse tomato and pepper farmers are likely to participate in a government grant program and to evaluate the program's impact on profitability. This study examines whether yield in kg, greenhouse value in euros, greenhouse area in square meters ($m^2$), distance to the market in kilometers (km), and other greenhouse crops grown influence the farmers' participation in a grant program. Genetic matching and propensity score matching (PSM) with logistic regression as the distance measure were methods utilized to match grantees and non-grantees. The second study with the use of these methods aimed to analyze the effect of the Ministry of Agriculture, Forestry and Rural Development (MAFRD) grant programs in the production of greenhouse tomatoes and peppers. A further research analysis included a comparison of the seasonal income among greenhouse tomato and pepper groups of farmers.

By researching the possibility of a greenhouse tomato and pepper farmer receiving a grant from MAFRD, the study's findings support the notion that farmers' participation in a grant program may be dependent on key covariates such as yield in kg, greenhouse value in euros, and greenhouse area in $m^2$. Policy researchers in Kosovo may also consider the evidence of the positive seasonal income difference of 1,777 euros found under genetic matching for the greenhouse tomato grantees compared to non-grantees. This evidence may help to identify farmers likely to be affected from the MAFRD grant programs. This study's findings could also be of interest to agencies for development that contribute to help MAFRD's efforts in Kosovo to provide new greenhouses.

The second study's results overall suggested that the quality and quantity of the matches obtained from the matching techniques were influenced by the sample size of the two groups of farmers. Although genetic matching has the tendency to provide better matches, it was not entirely conclusive that this method may be favored in all cases using this dataset over the method of PSM using logistic regression as the distance measure.

## References

Abdia, Y., K. Kulasekera, S. Datta, M. Boakye, and M. Kong. 2017. "Propensity Scores Based Methods for Estimating Average Treatment Effect and Average Treatment Effect Among Treated: A Comparative Study." *Biometrical Journal* 967-985.

Aigner, D. J., and S. F. Chu. 1968. "On Estimating the Industry Production Function." *American Economic Review* 826-839.

Aigner, D. J., C. A. Lovell, and P. Schmidt. 1977. "Formulation and estimation of stochastic frontier production functions." *Journal of Econometrics* 21-37.

Ali, A., M. Sharif, K. Mahmood, and N. Akmal. 2013. "Determinants of Cherry Production and Marketing in Pakistan: A Propensity Score Matching Approach." *Agricultural Economics Review*.

Archer, S. E. 2003. "Kosovo present and future." *Military Review* 31-40.

Arcotrass Gmbh; Vakakis International SA; Eurocare Gmbh;AKI. (2006). *Kosovo Report.* Arcotrass-Consortium.

Bajrami, E. 2016. "Evaluation of Agricultural Policy in the Dairy Sector in Kosovo and Efficiency Analysis at the Farm Level." MS Thesis, University of Arkansas.

Balk, B. M. 2001. "Scale Efficiency and Productivity Change." *Journal of Productivity Analysis* 159-183.

Balliu, A., and S. Kaçiu. 2008. *Potential of the Greenhouse Industry in Kosovo.* Horticultural Promotion in Kosovo (HPK).

Banker, R. D.,W. W. Cooper, and A. Charnes. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 1078-1092.

Bursac, Z., C. H. Gauss, D. K. Williams, and D. W. Hosmer. 2008. "Purposeful Selection of Variables in Logistic Regression." *Source Code for Biology and Medicine* 3-17.

Caliendo, M., and S. Kopeinig, 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 31-72.

Charles, V., and M. Kumar. 2012. *Data Envelopment Analysis and Its Applications to Management.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Charnes, A., W. Cooper, and E. Rhodes. 1978. *A Data Envelopment Analysis Approach to Evaluation of the Program Follow Through Experiment in U.S. Public School Education.* Pittsburgh, PA: Carnegie-Mellon University, School of Urban and Public Affairs.

Charnes, A., W. Cooper, and E. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Researc* 429-444.

Chivuraise, C., T. Chamboko, and G. Chagwiza. 2016. An Assessment of Factors Influencing Forest Harvesting in Smallholder Tobacco Production in Hurungwe District, Zimbabwe: An Application of Binary Logistic Regression Model. *Advances in Agriculture*.

Cooper, W. W., L. M. Seiford, and K. Tone. 2002. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Kluwer Academic Publishers.

Cox, D. R. 1958. "The Regression Analysis of Binary Sequences." *Journal of the Royal Statistical Society* 215-242.

D'Agostino, R. B. 1998. "Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-randomized Control Group." *Statistics in Medicine* 2265-2281.

Dehejia, R. H., and S. Wahba. 2002. "Propensity Score Matching Methods for Non-experimental Causal Studies." *Review of Economics and Statistics* 151–161.

Diamond, A., and J. S. Sekhon. 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* 932-945.

Elizabeth, P. 1981. "Why Turbulent Kosovo Has Marble Sidewalks but Troubled Industries." *The Christian Science Monitor*.

Fall, M., and T. Magnac. 2004. "How Valuable Is On-Farm Work to Farmers?" *American Journal of Agricultural Economics* 267-281.

Farrell, M. J. 1957. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society* 253-290.

Fried, H. O., K. C. Lovell, S. S. Schmidt. 2008. *The measurement of productive efficiency and productivity growth.* New York: Oxford University Press.

Gashi, P. 2017. "Free Trade and FDI in Kosovo: Prospects for Integration into the EU and Turkish Production Networks." *Turkish Economic Review* 86-95.

Graeub, B. E., M. J. Chappell, H. Wittman, S. Ledermann, R. B. Kerr, and B. Gemmill-Herren. 2016. "The State of Family Farms in the World." *World Development* 1-15.

Guo, S., and M. W. Fraser 2014. *Propensity Score Analysis: Statistical Methods and Applications.* SAGE Publications.

Hayes, A. F., and J. Matthes. 2009. "Computational Procedures for Probing Interactions in OLS and Logistic Regression: SPSS and SAS Implementations." *Behavior Research Methods* 924-36.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression.* New York: John Wiley and Sons, Incorporated.

International Monetary Fund. 2011. *Staff Report for the 2011 Article IV Consultation and the Initiation of a Staff-Monitored Program.* International Monetary Fund.

Johnes, J. 2006. "Data Envelopment Analysis and its Application to the Measurement of Efficiency in Higher Education." *Economics of Education Review* 273–288.

Jusufi, G., L. R. Mahmutaj, G. Jusufi, and N. Jusufi. 2015. "Kosovo's International Trade: Balance of Trade." *European Journal of Economics and Business Studies* 58-69.

Kaciu, S., I. Babaj, S. Aliu, and I. Demaj. 2016. "Potential of Protected Vegetable Production in Kosovo and Future Perspectives." *Acta Hortic* 461-466.

Korhonen, P., and T. Joro. 2015. *Extension of Data Envelopment Analysis with Preference Information: Value Efficiency.* Springer.

Kumar, S., and R. Gulati. 2008. "An Examination of Technical, Pure Technical, and Scale Efficiencies in Indian Public Sector Banks using Data Envelopment Analysis." *Eurasian Journal of Business and Economics* 33-69.

LaValley, M. P. 2008. "Logistic Regression." *Circulation* 2395-2399.

Legendre, A. M. 1805. *Nouvelles méthodes pour la détermination des orbites des comètes.* Firmin Didot.

Lever, J., M. Krzywinski, and N. Altman. 2016. "Points of Significance: Logistic Regression." *Nature Methods* 541-2.

Ministry of Agriculture, Forestry and Rural Development. 2016. *Green Report.* Prishtina: Ministry of Agriculture, Forestry and Rural Development.

McCaffrey, D. F., G. Ridgeway, and A. R. Morral. 2004. "Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies." *Psychol Methods* 403-25.

Mebane, W. R., and J. S. Sekhon. 1998. "Genetic Optimization Using Derivatives (GENOUD)". *Political Analysis* 187-210.

Meeusen, W., and J. Van Den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 435-444.

Ministry of Agriculture, Forestry and Rural Development. 2017. *Agriculture and Rural Development Program 2017.* Prishtina: Ministry of Agriculture, Forestry and Rural Development.

Reinhard, S., C. K. Lovell, and G. J. Thijssen. 2000. "Environmental Efficiency with Multiple Environmentally Detrimental Variables; Estimated with SFA and DEA." *European Journal of Operational Research* 287-303.

Rosenbaum, P. R., and D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 41-55.

Rubin, D. B. 1977. "Assignment to a Treatment Group on the Basis of a Covariate." *Journal of Educational Statistics* 1-26.

Sauer, J., M. Gorton, and S. Davidova. 2015. "Migration and Farm Technical Efficiency: Evidence from Kosovo." *Agricultural Economics* 629–641.

Stuart, E. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 1–21.

Stuart, E., and D. B. Rubin. 2008. *Best Practices in Quasi–Experimental Designs: Matching Methods for Causal Inference.* Thousand Oaks, CA: SAGE Publications Ltd.

Tanfani, E., and A. Testi. 2012. *Advanced Decision Making Methods Applied to Health Care.* Springer Science and Business Media.

Tolesa, G. N. 2017. “Logistic Regression Analysis of Marketability of Tomato Fruit Harvested at Different Maturity Stages and Subjected to Disinfection, Storage Condition and Storage Period Treatments.” *Biological Agriculture and Horticulture* 40-52.

Vuciterna, R. 2017. “Efficiency and Competitiveness of Kosovo Raspberry Producers.” *Selected Paper prepared for presentation at the Southern Agricultural Economics Association’s 2017 Annual Meeting*.

Westreich, D., J. Lessler, and M. J. Funk. 2010. “Propensity Score Estimation: Machine Learning and Classification Methods as Alternatives to Logistic Regression.” *Journal of Clinical Epidemiology* 826–833.

## Appendix

INSTITUTIONAL REVIEW BOARD (IRB) #17-04-678 PROTOCOL APPROVAL

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

May 16, 2017

MEMORANDUM

| | |
|---|---|
| TO: | Blend Frangu<br>Jennie Popp |
| FROM: | Ro Windwalker<br>IRB Coordinator |
| RE: | New Protocol Approval |
| IRB Protocol #: | 17-04-678 |
| Protocol Title: | Tomato and Banana Pepper Production in Kosovo |
| Review Type: | ☒ EXEMPT ☐ EXPEDITED ☐ FULL IRB |
| Approved Project Period: | Start Date: 05/15/2017 Expiration Date: 05/14/2018 |

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form Continuing Review for IRB Approved Projects, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (https://vpred.uark.edu/units/rscp/index.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

**This protocol has been approved for 400 participants.** If you wish to make any modifications in the approved protocol, including enrolling more than this number, you must seek approval prior to implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

109 MLKG • 1 University of Arkansas • Fayetteville, AR 72701-1201 • (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu
The University of Arkansas is an equal opportunity/affirmative action institution.

SURVEY A: GREENHOUSE TOMATO PRODUCTION IN KOSOVO, 2017

## GENERAL INFORMATION

This research survey is conducted by Blend Frangu under Agricultural Economics and Agribusiness | University of Arkansas

**Survey purpose**

The purpose of this survey is to obtain information on the production of greenhouse fresh market tomatoes and peppers in the Republic of Kosovo. Your information may be used for thesis research purposes. The research results of the tomato and pepper input efficiency will present helpful guidance for Kosovar farmers. The results may also be published as a master's thesis and/or in academic journals.

**Research study benefits**

The act of participating in this survey will not provide direct benefits. Nevertheless, your responses may help us determine factors that impact the financial health of greenhouse pepper and tomato farms. The results of the study may help you improve your greenhouse tomato and pepper input efficiency.

**Survey risks**

There is no anticipated risk, sensitive question or discomfort for farmers when completing the survey.

**Data management**

The collection of the information and your answers are recorded anonymously. The research analysis does not identify any farmer by name. Your participation in this research study is voluntary. You may choose not to participate. If you see fit to participate in this research survey, you may withdraw your consent at any time and no one will know whether you participated in the research study.

**Contact**

If you have any questions about the research study procedures or survey questions, you may contact the principal researcher Blend Frangu via electronic mail at bfrangu@uark.edu or the research supervisor Jennie Popp, Ph.D. at jhpopp@uark.edu. For questions or concerns about your rights as a research participant, please contact Ro Windwalker, the University's Compliance Coordinator, at (479)-575-2208 or by electronic mail at irb@uark.edu.

Please tick your choice below. You may want a copy of this consent form for your reference. Ticking on the "Agree" box indicates that

I. You have read and understood the above information
II. You voluntarily agree to participate in the research survey
III. You are 18 years of age or older

☐ Agree

☐ Disagree

SURVEY A: GREENHOUSE PRODUCTION IN KOSOVO, 2017

## MODULE 1. MAIN RESPONDENT

| | |
|---|---|
| No. of survey: | |
| Date: | ___/____/__________ |
| Highest qualification: | |
| Education (in years): | |
| Gender of the farmer: Male/Female | |
| Interviewee age: | |
| Number of family members: | |
| Greenhouse production experience (in years): | |
| Village: | |
| Municipality: | |
| Region: | |

## MODULE 2. GREENHOUSE DATA

1. What is your greenhouse type?
   (Please check one or more)

   ☐ Glass covering ☐ Polyethylene covering
   ☐ Plastic tunnel ☐ Rigid plastic tunnel

2. What are the dimensions of your greenhouse structure?

   Glass covering Length _______m Width _______m
   Plastic tunnel Length _______m Width _______m
   Polyethylene covering Length _______m Width _______m
   Rigid plastic tunnel Length _______m Width _______m

3. What is your total greenhouse area? __________________$m^2$

4. How much did your greenhouse cost to build? __________________(€)

| Heating, cooling, irrigation systems and water usage questions | Options | | |
|---|---|---|---|
| | Yes | No | NA |
| 5. Do you have a drip irrigation system? | ☐ Yes; | ☐ No; | ☐ NA; |
| 6. If no, please specify your alternative irrigation system | | | |
| 7. Do you conduct farm water testing? | ☐ Yes; | ☐ No; | ☐ NA; |
| 8. If yes, what is the cost per test in euros? | | | |
| 9. Do you conduct soil testing? | ☐ Yes; | ☐ No; | ☐ NA; |
| 10. If yes, what is the cost per test in euros? | | | |
| 11. Do you have a ventilation system? | ☐ Yes; | ☐ No; | ☐ NA; |
| 12. If yes, what type and how much did it cost in euros? | | | |
| 13. Do you have a heating system? | ☐ Yes; | ☐ No; | ☐ NA; |
| 14. If yes, what type and how much did it cost in euros? | | | |
| 15. What is the power source of your irrigation pump? | | | |
| 16. What is the depth to ground water at the well site in meters? | | | |

| 17. What is your capital investment for greenhouse production? | |
|---|---|
| Investment category | Total (€) |
| Land - Building site | |
| Greenhouse building | |
| Irrigation system | |
| Farm machinery and equipment | |
| If Other, please specify: | |

18. How many months was your greenhouse in operation in the last two years?

In 2016 ___________________
In 2015 ___________________

19. Did you receive an investment grant for your greenhouse from Ministry of Agriculture, Forestry and Rural Development?

☐ Yes ☐ No

20. If yes, what was its monetary value? ___________________(€)

21. Do you practice vegetable crop rotation? ☐ Yes ☐ No

22. Which vegetable crops did you grow in 2016?
(Please check one or more)

☐ Tomatoes ☐ Peppers
☐ Lettuce ☐ Cucumbers
☐ Spinach ☐ Eggplants
☐ If Other, please specify: ___________________________________________________

23. Which vegetable crops do you reason are more profitable to grow in the greenhouse?
(Please check one or more)

☐ Tomatoes ☐ Peppers
☐ Lettuce ☐ Cucumbers
☐ If Other, please specify: ________________________________________________

## MODULE 3. GREENHOUSE TOMATO FARM DATA

1. In what year did you begin greenhouse tomato production? ___________________
2. What was your greenhouse tomato production area in 2016? ___________________ $m^2$
3. Which tomato varieties and/or cultivars do you grow?

Variety 1 ___________________ Cultivar 1 ___________________ _____% share
Variety 2 ___________________ Cultivar 2 ___________________ _____% share
Variety 3 ___________________ Cultivar 3 ___________________ _____% share

4. What are your greenhouse tomato plant spacings?

In row ___________________cm
Between row ___________________cm
Rows per house ___________________

5. Do you use trellising? ☐ Yes ☐ No

6. If yes, which type of trellising do you use?

☐ Basketweave system ☐ Hanging string system

If Other, please specify: ______________________________________________

| 7. What is your greenhouse tomato labor use? | Options | | | | | | |
|---|---|---|---|---|---|---|---|
| | Working days/year | | Wage rate/hour (€) | | Salary/month (€) | | |
| / | $L_F$[A] | $L_H$[B] | $L_F$ | $L_H$ | $L_F$ | $L_H$ | NA |
| Plotting | | | | | | | |
| Seeding | | | | | | | |
| Watering | | | | | | | |
| Fertilizer and pesticide | | | | | | | |
| Harvesting | | | | | | | |
| Trellising | | | | | | | |
| Pruning | | | | | | | |
| / | | | | Total: | | | / |

[A] $L_F$ as family labor corresponds to labor coming from family members
[B] $L_H$ as hired labor corresponds to labor coming from non-family members

| 8. What is your fertilizer use? | Type | Amount | Cost (€) | NA |
|---|---|---|---|---|
| Planting | | | | |
| Flowering - fruit-set | | | | |
| Harvesting | | | | |
| 9. What is your pesticide use? | Type[C] | Amount | Cost (€) | NA |
| Herbicide | | | | |
| Insecticide | | | | |
| Fungicide | | | | |
| Larvicides | | | | |
| 10. How much farm water per square meter do you apply each day? | | | | |

[D] In respect to the type of farm water applied, please indicate if it is drinking and/or non-drinking water

11. How warm do you keep your tomato indoor growing area? _________________ °C

12. What was your tomato total yield in 2016? _________________kg/area

13. Which is your primary marketing channel to sell your greenhouse tomato produce? (Please check and fill one or more)

☐ Direct-to-consumer - Price ______(€) ☐ Restaurants - Price ______(€)
☐ Grocery/retail stores - Price ______(€) ☐ Farmer’s market - Price ______(€)
☐ If Other, please specify: ______________________________________________

14. What is your distance to the market? _________________km

15. What is your quantity and selling price per kilogram of tomatoes in the harvesting season?

Early-season __________________ (€) Quantity __________________ kg
Mid-season __________________ (€) Quantity __________________ kg
Late-season __________________ (€) Quantity __________________ kg

16. Do you follow Good Agricultural Practices (GAP) for tomato production?

☐ Yes ☐ No

17. Do you have external income outside greenhouse tomato production?

☐ Yes ☐ No

18. If yes, what is your annual income amount? __________________(€)

19. Do you have external income from remittances?

☐ Yes ☐ No

20. If yes, what is your annual income amount? __________________(€)

21. How many days per week do you irrigate tomatoes? __________________days

22. How many minutes per day do you irrigate tomatoes? __________________minutes

23. What is your irrigation water source? __________________

## MODULE 4. GREENHOUSE PRODUCER NEEDS

1. For each of the following statements below, please assign a number to each of them as the response that best characterizes your level of agreement or disagreement, where:

| Table 1. Numbers 1-5 representing your level of agreement or disagreement to each statement | | | | |
|---|---|---|---|---|
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| 1 | 2 | 3 | 4 | 5 |

As a vegetable farmer I need:

A. _______ output price support
B. _______ input subsidy
C. _______ an investment grant for greenhouse structure upgrade
D. _______ an investment grant for an enhanced heating system
E. _______ an investment grant for an enhanced cooling system
F. _______ an investment grant for an enhanced ventilation system
G. _______ improved irrigation infrastructure
H. _______ greenhouse automation training
I. _______ wastewater management training
J. _______ nutritional management training
K. _______ plant growth regulators training
L. _______ greenhouse cooling and/or heating management training

SURVEY B: GREENHOUSE PEPPER PRODUCTION IN KOSOVO, 2017

| GENERAL INFORMATION |
|---|
| This research survey is conducted by Blend Frangu under Agricultural Economics and Agribusiness \| University of Arkansas |

**Survey purpose**

The purpose of this survey is to obtain information on the production of greenhouse fresh market tomatoes and peppers in the Republic of Kosovo. Your information may be used for thesis research purposes. The research results of the tomato and pepper input efficiency will present helpful guidance for Kosovar farmers. The results may also be published as a master's thesis and/or in academic journals.

**Research study benefits**

The act of participating in this survey will not provide direct benefits. Nevertheless, your responses may help us determine factors that impact the financial health of greenhouse pepper and tomato farms. The results of the study may help you improve your greenhouse tomato and pepper input efficiency.

**Survey risks**

There is no anticipated risk, sensitive question or discomfort for farmers when completing the survey.

**Data management**

The collection of the information and your answers are recorded anonymously. The research analysis does not identify any farmer by name. Your participation in this research study is voluntary. You may choose not to participate. If you see fit to participate in this research survey, you may withdraw your consent at any time and no one will know whether you participated in the research study.

**Contact**

If you have any questions about the research study procedures or survey questions, you may contact the principal researcher Blend Frangu via electronic mail at bfrangu@uark.edu or the research supervisor Jennie Popp, Ph.D. at jhpopp@uark.edu. For questions or concerns about your rights as a research participant, please contact Ro Windwalker, the University's Compliance Coordinator, at (479)-575-2208 or by electronic mail at irb@uark.edu.

Please tick your choice below. You may want a copy of this consent form for your reference. Ticking on the "Agree" box indicates that

IV. You have read and understood the above information
V. You voluntarily agree to participate in the research survey
VI. You are 18 years of age or older

☐ Agree

☐ Disagree

SURVEY B: GREENHOUSE PRODUCTION IN KOSOVO, 2017

## MODULE 1. MAIN RESPONDENT

| | |
|---|---|
| No. of survey: | |
| Date: | ___/____/__________ |
| Highest qualification: | |
| Education (in years): | |
| Gender of the farmer: Male/Female | |
| Interviewee age: | |
| Number of family members: | |
| Greenhouse production experience (in years): | |
| Village: | |
| Municipality: | |
| Region: | |

## MODULE 2. GREENHOUSE DATA

24. What is your greenhouse type?
(Please check one or more)

☐ Glass covering ☐ Polyethylene covering
☐ Plastic tunnel ☐ Rigid plastic tunnel

25. What are the dimensions of your greenhouse structure?

| | | |
|---|---|---|
| Glass covering | Length _______m | Width _______m |
| Plastic tunnel | Length _______m | Width _______m |
| Polyethylene covering | Length _______m | Width _______m |
| Rigid plastic tunnel | Length _______m | Width _______m |

26. What is your total greenhouse area? _________________$m^2$

27. How much did your greenhouse cost to build? _________________(€)

| Heating, cooling, irrigation systems and water usage questions | Options | | |
|---|---|---|---|
| | Yes | No | NA |
| 28. Do you have a drip irrigation system? | ☐ Yes; | ☐ No; | ☐ NA; |
| 29. If no, please specify your alternative irrigation system | | | |
| 30. Do you conduct farm water testing? | ☐ Yes; | ☐ No; | ☐ NA; |
| 31. If yes, what is the cost per test in euros? | | | |
| 32. Do you conduct soil testing? | ☐ Yes; | ☐ No; | ☐ NA; |
| 33. If yes, what is the cost per test in euros? | | | |
| 34. Do you have a ventilation system? | ☐ Yes; | ☐ No; | ☐ NA; |
| 35. If yes, what type and how much did it cost in euros? | | | |
| 36. Do you have a heating system? | ☐ Yes; | ☐ No; | ☐ NA; |
| 37. If yes, what type and how much did it cost in euros? | | | |

| | |
|---|---|
| 38. What is the power source of your irrigation pump? | |
| 39. What is the depth to ground water at the well site in meters? | |

| 40. What is your capital investment for greenhouse production? | |
|---|---|
| Investment category | Total (€) |
| Land - Building site | |
| Greenhouse building | |
| Irrigation system | |
| Farm machinery and equipment | |
| If Other, please specify: | |

41. How many months was your greenhouse in operation in the last two years?

In 2016 ________________

In 2015 ________________

42. Did you receive an investment grant for your greenhouse from Ministry of Agriculture, Forestry and Rural Development?

☐ Yes ☐ No

43. If yes, what was its monetary value? ________________(€)

44. Do you practice vegetable crop rotation? ☐ Yes ☐ No

45. Which vegetable crops did you grow in 2016?
(Please check one or more)

☐ Tomatoes ☐ Peppers
☐ Lettuce ☐ Cucumbers
☐ Spinach ☐ Eggplants
☐ If Other, please specify: ________________________________________

46. Which vegetable crops do you reason are more profitable to grow in the greenhouse?
(Please check one or more)

☐ Tomatoes ☐ Peppers
☐ Lettuce ☐ Cucumbers
☐ If Other, please specify: ________________________________________

## MODULE 3. GREENHOUSE PEPPER FARM DATA

1. In what year did you begin greenhouse pepper production? ________________

2. Do you use trellising? ☐ Yes ☐ No

3. If yes, which type of trellising do you use?

☐ Basketweave system ☐ Hanging string system

If Other, please specify: ________________________________________

| 4. What is your greenhouse pepper labor use? | Options | | | | | | |
|---|---|---|---|---|---|---|---|
| | Working days/year | | Wage rate/hour (€) | | Salary/month (€) | | |
| / | $L_F$[D] | $L_H$[E] | $L_F$ | $L_H$ | $L_F$ | $L_H$ | NA |
| Plotting | | | | | | | |
| Seeding | | | | | | | |
| Watering | | | | | | | |
| Fertilizer and pesticide | | | | | | | |
| Harvesting | | | | | | | |
| Trellising | | | | | | | |
| Pruning | | | | | | | |
| / | | | | Total: | | | / |

[D] $L_F$ as family labor corresponds to labor coming from family members
[E] $L_H$ as hired labor corresponds to labor coming from non-family members

| 5. What is your fertilizer use? | Type | Amount | Cost (€) | NA |
|---|---|---|---|---|
| Planting | | | | |
| Flowering - fruit-set | | | | |
| Harvesting | | | | |
| 6. What is your pesticide use? | Type[F] | Amount | Cost (€) | NA |
| Herbicide | | | | |
| Insecticide | | | | |
| Fungicide | | | | |
| Larvicides | | | | |
| 7. How much farm water per square meter do you apply each day? | | | | |

[F] In respect to the type of farm water applied, please indicate if it is drinking and/or non-drinking water

8. Which pepper varieties and/or cultivars do you grow?

Variety 1 __________________ Cultivar 1 __________________ ____% share
Variety 2 __________________ Cultivar 2 __________________ ____% share
Variety 3 __________________ Cultivar 3 __________________ ____% share

9. How warm do you keep your pepper indoor growing area? _________________ $^{o}C$

10. What was your greenhouse pepper production area in 2016? _______________ $m^2$

11. What was your pepper total yield in 2016? __________________kg/area

12. What are your greenhouse pepper plant spacings?

In row __________________cm

Between row __________________cm

Rows per house __________________

13. Which is your primary marketing channel to sell your greenhouse pepper produce? (Please check and fill one or more)

- ☐ Direct-to-consumer - Price ______(€)
- ☐ Restaurants - Price ______(€)
- ☐ Grocery/retail stores - Price ______(€)
- ☐ Farmer's market - Price ______(€)
- ☐ If Other, please specify: ________________________________________________

14. What is your distance to the market? _________________km
15. What is your quantity and selling price per kilogram of peppers in the harvesting season?

Early-season _________________ (€) Quantity _________________ kg
Mid-season _________________ (€) Quantity _________________ kg
Late-season _________________ (€) Quantity _________________ kg

16. Do you follow Good Agricultural Practices (GAP) for pepper production?

☐ Yes ☐ No

17. Do you have external income outside greenhouse pepper production?

☐ Yes ☐ No

18. If yes, what is your annual income amount? _________________(€)
19. Do you have external income from remittances?

☐ Yes ☐ No

20. If yes, what is your annual income amount? _________________(€)
21. How many days per week do you irrigate peppers? _________________days
22. How many minutes per day do you irrigate peppers? _________________minutes
23. What is your irrigation water source? _________________

## MODULE 4. GREENHOUSE PRODUCER NEEDS

2. For each of the following statements below, please assign a number to each of them as the response that best characterizes your level of agreement or disagreement, where:

| Table 1. Numbers 1-5 representing your level of agreement or disagreement to each statement | | | | |
|---|---|---|---|---|
| Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| 1 | 2 | 3 | 4 | 5 |

As a vegetable farmer I need:

A. _______ output price support
B. _______ input subsidy
C. _______ an investment grant for greenhouse structure upgrade
D. _______ an investment grant for an enhanced heating system
E. _______ an investment grant for an enhanced cooling system
F. _______ an investment grant for an enhanced ventilation system
G. _______ improved irrigation infrastructure
H. _______ greenhouse automation training
I. _______ wastewater management training
J. _______ nutritional management training
K. _______ plant growth regulators training
L. _______ greenhouse cooling and/or heating management training

R PROGRAMMING SCRIPT:

EVALUATING GREENHOUSE TOMATO AND PEPPER INPUT EFFICIENCY USE IN KOSOVO

```
 "
     R Programming Script
     Study:  Evaluating Greenhouse Tomato and Pepper Input Efficiency Use in Kosovo
     Author: Blend Frangu - Graduate student at the University of Arkansas,
             the Department of Agricultural Economics and Agribusiness.
     "
     # Reading the csv data file and subsetting data------------------------------

     data<-read.csv(file.choose(),header=T)
     tomato<-data[1:94,]
     pepper<-data[95:136,]

     # Defining new variable names------------------------------------------

     "Greenhouse tomatoes"

     "Data envelopment analysis variables"
     tomato$THOUSEVAL<-tomato$greenhousevalue
     tomato$TOFERT<-tomato$porganicfertilizer
     tomato$TAFERT<-tomato$partificialfertilizer
     tomato$TCFERT<-tomato$fcrystallinefertilizer
     tomato$TEAFERT<-tomato$fartificialfertilizer
     tomato$TINSEC<-tomato$insecticideliters
     tomato$TAREA<-tomato$tompeparea
     tomato$TYIELD<-tomato$yield
     tomato$TLABOR<-(tomato$lfplotdays+tomato$lfseeddays+tomato$lfchemicaldays+
                     tomato$lfharvestdays+tomato$lftrellisingdays+tomato$lfpruningdays+
                     tomato$lfwaterdays+tomato$lhwaterdays+tomato$lhchemicaldays+
                     tomato$lhharvestdays+tomato$lhplotdays+tomato$lhseeddays+
                     tomato$lhtrellisingdays+tomato$lhpruningdays)

     "Linear and logistic regression variables"
     tomato$TCROP<-tomato$nutrition
     tomato$TPOWER<-tomato$power
     tomato$TROWS<-tomato$rowshouse
     tomato$TWHOLE<-tomato$wholesaleprice
     tomato$TWATER<-tomato$irrigationvalue
     tomato$TEDU<-tomato$eduyears
     tomato$TFAMILY<-tomato$family
     tomato$TEXREV<-tomato$exrevenue
     tomato$TFARM<-tomato$farmermarketprice
```

```
tomato$TOTHER<-tomato$combinationcrops
tomato$TWELL<-tomato$welldepth

"Greenhouse peppers"

"Data envelopment analysis variables"
pepper$PHOUSEVAL<-pepper$greenhousevalue
pepper$POFERT<-pepper$porganicfertilizer
pepper$PCFERT<-pepper$fcrystallinefertilizer
pepper$PEAFERT<-pepper$fartificialfertilizer
pepper$PINSEC<-pepper$insecticideliters
pepper$PAREA<-pepper$tompeparea
pepper$PYIELD<-pepper$yield
pepper$PLABOR<-(pepper$lfplotdays+pepper$lfseeddays+pepper$lfwaterdays+
              pepper$lfchemicaldays+pepper$lfharvestdays+pepper$lftrellisingdays+
              pepper$lfpruningdays+pepper$lhplotdays+pepper$lhseeddays+
              pepper$lhwaterdays+pepper$lhchemicaldays+pepper$lhharvestdays+
              pepper$lhtrellisingdays+pepper$lhpruningdays)

"Linear and logistic regression variables"
pepper$PEXREV<-pepper$exrevenue
pepper$PFARM<-pepper$farmermarketprice
pepper$POTHER<-pepper$combinationcrops
pepper$PWELL<-pepper$welldepth
pepper$PWATER<-pepper$irrigationvalue
pepper$PEDU<-pepper$eduyears
pepper$PFAMILY<-pepper$family
pepper$PCROP<-pepper$nutrition
pepper$PPOWER<-pepper$power
pepper$PROWS<-pepper$rowshouse
pepper$PWHOLE<-pepper$wholesaleprice

# Results: Greenhouse tomatoes----------------------------------------

"Inputs and output"
x1<-with(tomato,cbind(TAREA,THOUSEVAL,TLABOR,TINSEC,TOFERT,TAFERT,TCFERT,
         TEAFERT))
y1<-with(tomato,TYIELD)

"Greenhouse tomato efficiency scores under different returns to scale"
vrsdea<-Benchmarking::dea(x1,y1,RTS="vrs",ORIENTATION="in")
vrsdea
summary(vrsdea)
crsdea<-Benchmarking::dea(x1,y1,RTS="crs",ORIENTATION="in")
crsdea
summary(crsdea)
```

```
drsdea<-Benchmarking::dea(x1,y1,RTS="drs",ORIENTATION="in")
drsdea
summary(drsdea)

"Scale efficiency (SE) analysis"
scale.effscore<-Benchmarking::eff(crsdea)/Benchmarking::eff(vrsdea)
summary(scale.effscore)
scale.efforientation<-tomato$SEtype<-ifelse(scale.effscore==1,"SE",
         ifelse(Benchmarking::eff(crsdea)==Benchmarking::eff(drsdea),"IRS","DRS"))
dearesult<-as.data.frame(cbind(round(Benchmarking::eff(vrsdea),10),
                               round(Benchmarking::eff(crsdea),10),
                               round(Benchmarking::eff(drsdea),10),
                               round(scale.effscore,10),scale.efforientation))
ww<-cbind(dearesult,tomato$region)
colnames(ww)<-c("VRS","CRS","DRS","SE","scale","region")
ww

"Additional new variables"
tomato$scale.effscore<-scale.effscore
ww$TYIELD<-tomato$TYIELD

"Efficiency scores and returns to scale (RTS) result frequencies"

"Variable returns to scale (VRS) efficiency score frequencies by region"
table(ww$VRS,tomato$region)
prop.table(table(ww$VRS,tomato$region))

"Constant returns to scale (CRS) efficiency score frequencies by region"
table(ww$CRS,tomato$region)
prop.table(table(ww$CRS,tomato$region))

"Scale efficiency (SE) score frequencies by region"
table(ww$SE,tomato$region)
prop.table(table(ww$SE,tomato$region))

"BCC and CCR models: Mean, min, max, and standard deviation of efficiency scores by
region"
tomato$vrsdea<-vrsdea$eff
tomato$crsdea<-crsdea$eff
aggregate(vrsdea~region,FUN=mean,data=tomato)
aggregate(crsdea~region,FUN=mean,data=tomato)
aggregate(vrsdea~region,FUN=min,data=tomato)
aggregate(crsdea~region,FUN=min,data=tomato)
aggregate(vrsdea~region,FUN=max,data=tomato)
aggregate(crsdea~region,FUN=max,data=tomato)
```

```r
aggregate(vrsdea~region,FUN=sd,data=tomato)
aggregate(crsdea~region,FUN=sd,data=tomato)
aggregate(scale.effscore~region,FUN=mean,data=tomato)
aggregate(scale.effscore~region,FUN=min,data=tomato)
aggregate(scale.effscore~region,FUN=max,data=tomato)
aggregate(scale.effscore~region,FUN=sd,data=tomato)

"Aggregate function: Finding input and output means according to the returns to
scale (RTS)"
aggregate(TYIELD~SEtype,FUN=mean,data=tomato)
aggregate(TYIELD~SEtype,FUN=sd,data=tomato)
aggregate(TAREA~SEtype,FUN=mean,data=tomato)
aggregate(TAREA~SEtype,FUN=sd,data=tomato)
aggregate(TINSEC~SEtype,FUN=mean,data=tomato)
aggregate(TINSEC~SEtype,FUN=sd,data=tomato)
aggregate(TOFERT~SEtype,FUN=mean,data=tomato)
aggregate(TOFERT~SEtype,FUN=sd,data=tomato)
aggregate(TAFERT~SEtype,FUN=mean,data=tomato)
aggregate(TAFERT~SEtype,FUN=sd,data=tomato)
aggregate(TEAFERT~SEtype,FUN=mean,data=tomato)
aggregate(TEAFERT~SEtype,FUN=sd,data=tomato)
aggregate(TCFERT~SEtype,FUN=mean,data=tomato)
aggregate(TCFERT~SEtype,FUN=sd,data=tomato)
aggregate(TLABOR~SEtype,FUN=mean,data=tomato)
aggregate(TLABOR~SEtype,FUN=sd,data=tomato)
aggregate(THOUSEVAL~SEtype,FUN=mean,data=tomato)
aggregate(THOUSEVAL~SEtype,FUN=sd,data=tomato)

"The greenhouse tomato multiple binary logistic regression"

"Explanatory variable selection: The exhaustive search procedure for greenhouse
tomatoes"
tomato$y<-ifelse(crsdea$eff>=1,1,0)
library(bestglm)
res.bestglm<-
subset(tomato,select=c("family","grant","rank","distance","nutritionalthelp",
                       "welldepth","lfwaterdays","wholesaleprice",
                       "groceryprice","otherp","power","EQ","rowshouse",
                       "eduyears","age","expyears","gender","soiltest",
                       "polyethylene","gutter","combinationcrops",
                       "irrigationvalue","croprotation","equipmentvalue",
                       "nutrition","y"))
model<-bestglm(Xy=res.bestglm,family=binomial,IC="AIC",method="exhaustive")

"The choice of 5 best models"
model$BestModels
```

```r
"The selected model"
summary(model$BestModel)

"Estimating greenhouse tomato multiple binary logistic regression"

"The multiple logistic regression model"
tomato$crsdea<-crsdea$eff
tomato$TOUT<-ifelse(crsdea$eff>=1,1,0)
logitt<-glm(TOUT~TFAMILY+TEDU+TWATER+TCROP+TPOWER+TROWS+TWHOLE,
            data=tomato,family="binomial")
summary(logitt)

"Logistic regression estimates in odds ratios"
library(oddsratio)
or_glm(data=data_glm,model=logitt)

"Logistic regression Nagelkerke R-squared"
library(rms)
lrm(TOUT~TFAMILY+TEDU+TWATER+TCROP+TPOWER+TROWS+TWHOLE,data=tomato)

"Plots for the greenhouse tomato logistic regression estimates in odds ratios"
library(sjPlot)
library(sjmisc)
modelplot<-plot_model(logitt,axisLabels.y=lab,transformTicks=F,sort.est=T,
                      facet.grid=F, colors=c("gray40","black"),show.values=T,
                      value.offset=0.4,digits=4,value.size=3,dot.size=1,
                      type="est",line.size=.3,show.data=T,
                      axis.labels=c("Wholesaleprice/kg",
                      "Power source (electricity/fuel)",
                      "Crop nutrition training","Education in years",
                      "Irrigation in euro value",
                      "Family members","Rows/greenhouse"),
                      title="Greenhouse Tomato Logistic Estimate Ranking",
                      axis.title="Odds Ratios",width=0.3)
modelplot<-modelplot+theme_set(theme_bw())
modelplot<-modelplot+theme(panel.grid.major=element_blank(),
                           panel.grid.minor=element_blank())
modelplot<-modelplot+theme(text=element_text(family="Times",size=8.5))
modelplot

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(modelplot)
dev.off()

"The greenhouse tomato multiple linear regression analysis"
```

```r
"Explanatory variable selection: The exhaustive search procedure for greenhouse
tomatoes"
library(leaps)
tomato$crsdea<-crsdea$eff
olsmodel<-regsubsets(crsdea~family+grant+grantvalue+rank+distance+nutrition+
                     lfwaterdays+combinationcrops+otherp+power+EQ+rowshouse+
                     irrigationvalue+croprotation+equipmentvalue+polyethylene+
                     gutter+eduyears+age+expyears+gender+soiltest+X2016usage+
                     wholesale+wholesaleprice+consumerprice,data=tomato,nbest=1,
                     nvmax=NULL,force.in=NULL,force.out=NULL,method="exhaustive")
olsmodel
summary.out<-summary(olsmodel)
as.data.frame(summary.out$outmat)
plot(olsmodel,scale="adjr2")
title("Tomato Regressors based on Adjusted R-squared",adj=0.04,line=0.7,
      font.main=6,cex.main=1.4)

"The multiple linear regression model"
tomato$TDEP<-crsdea$eff
lineart<-lm(TDEP~TFAMILY+TEDU+TWATER+TCROP+TPOWER+TROWS+TWHOLE,data=tomato)
summary(lineart)

"Plots for the greenhouse tomato linear regression estimates"
modelplot1<-plot_model(lineart,axisLabels.y=lab,transformTicks=F,sort.est=T,
                       facet.grid=F, colors=c("gray40","black"),show.values=T,
                       value.offset=0.4,digits=4, value.size=3,dot.size=1,
                       type="est",line.size=0.3,show.data=T,
                       axis.labels=c("Wholesaleprice/kg",
                       "Power source (electricity/fuel)",
                       "Crop nutrition training","Education in years",
                       "Irrigation in euro value","Family members",
                       "Rows/greenhouse"),
                       title="Greenhouse Tomato Linear Estimate Ranking",
                       axis.title="Estimate Values",width=0.3)
modelplot1<-modelplot1+theme_set(theme_bw())
modelplot1<-modelplot1+theme(panel.grid.major=element_blank(),
                             panel.grid.minor=element_blank())
modelplot1<-modelplot1+theme(text=element_text(family="Times",size=8.5))
modelplot1

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(modelplot1)
dev.off()
```

```
"Summary of the greenhouse tomato variable descriptive statistics for data
envelopment analysis"
variables<-with(tomato,cbind(TAREA,THOUSEVAL,TLABOR,TINSEC,TOFERT,TAFERT,TCFERT,
                TEAFERT,TYIELD))
library(stargazer)
stargazer(list(variables),type="text",summary.logical=T,single.row=T,
          font.size="small",median=T,min.max=T,summary=T,digits=0)

"Coefficients of variation for the greenhouse tomato data envelopment analysis
model"
sd(tomato$TINSEC)/mean(tomato$TINSEC)
sd(tomato$TLABOR)/mean(tomato$TLABOR)
sd(tomato$THOUSEVAL)/mean(tomato$THOUSEVAL)
sd(tomato$TAREA)/mean(tomato$TAREA)
sd(tomato$TOFERT)/mean(tomato$TOFERT)
sd(tomato$TAFERT)/mean(tomato$TAFERT)
sd(tomato$TCFERT)/mean(tomato$TCFERT)
sd(tomato$TEAFERT)/mean(tomato$TEAFERT)
sd(tomato$TYIELD)/mean(tomato$TYIELD)

"Summary of the greenhouse tomato variable descriptive statistics for the
linear/logistic model"
variablesloglin<-with(tomato,cbind(TCROP,TPOWER,TROWS,TWHOLE,TWATER,
                      TEDU,TFAMILY,TEXREV,TFARM,TOTHER,TWELL,TDEP,TOUT))
stargazer(list(variablesloglin),type="text",summary.logical=T,single.row=T,
          font.size="small",median=T,min.max=T,summary=T)

"Coefficients of variation for the greenhouse tomato linear/logistic model"
sd(tomato$TCROP)/mean(tomato$TCROP)
sd(tomato$TPOWER)/mean(tomato$TPOWER)
sd(tomato$TROWS)/mean(tomato$TROWS)
sd(tomato$TWHOLE)/mean(tomato$TWHOLE)
sd(tomato$TWATER)/mean(tomato$TWATER)
sd(tomato$TEDU)/mean(tomato$TEDU)
sd(tomato$TFAMILY)/mean(tomato$TFAMILY)
sd(tomato$TDEP)/mean(tomato$TDEP)
sd(tomato$TOUT)/mean(tomato$TOUT)
sd(tomato$TEXREV,na.rm=T)/mean(tomato$TEXREV,na.rm=T)
sd(tomato$TFARM)/mean(tomato$TFARM)
sd(tomato$TOTHER)/mean(tomato$TOTHER)
sd(tomato$TWELL)/mean(tomato$TWELL)

# Results: Greenhouse peppers-------------------------------------

"Inputs and output"
x2<-with(pepper,cbind(PAREA,PLABOR,POFERT,PINSEC,PEAFERT,PCFERT))
```

```r
y2<-with(pepper,PYIELD)

"Greenhouse pepper efficiency scores under different returns to scale (RTS)"
vrsdea1<-Benchmarking::dea(x2,y2,RTS="vrs",ORIENTATION="in")
vrsdea1
summary(vrsdea1)
crsdea1<-Benchmarking::dea(x2,y2,RTS="crs",ORIENTATION="in")
crsdea1
summary(crsdea1)
drsdea1<-Benchmarking::dea(x2,y2,RTS="drs",ORIENTATION="in")
drsdea1
summary(drsdea1)

"Scale efficiency (SE) analysis"
scale.effscore1<-Benchmarking::eff(crsdea1)/Benchmarking::eff(vrsdea1)
scale.effscore1
summary(scale.effscore1)
scale.efforientation1<-pepper$SEtype1<-ifelse(scale.effscore1==1,"Scale Efficient",
        ifelse(Benchmarking::eff(crsdea1)==Benchmarking::eff(drsdea1),"IRS","DRS"))
dearesult1<-as.data.frame(cbind(round(Benchmarking::eff(vrsdea1),20),
                                round(Benchmarking::eff(crsdea1),20),
                                round(Benchmarking::eff(drsdea1),20),
                                round(scale.effscore1,20),scale.efforientation1))
ww1<-cbind(dearesult1,pepper$region)
colnames(ww1)<-c("VRS","CRS","DRS","SE","scale","region")
ww1

"Additional new variable"
pepper$scale.effscore<-scale.effscore1

"Efficiency scores and returns to scale (RTS) result frequencies"

"Variable returns to scale (VRS) efficiency score frequencies by region"
table(ww1$VRS,ww1$region)
prop.table(table(ww1$VRS,ww1$region))

"Constant returns to scale (CRS) efficiency score frequencies by region"
table(ww1$CRS,ww1$region)
prop.table(table(ww1$CRS,ww1$region))
summary(crsdea1)
"Scale efficiency (SE) score frequencies by region"
table(ww1$scale,ww1$region)
prop.table(table(ww1$scale,ww1$region))

"BCC and CCR models: Mean, min, max, and standard deviation of efficiency scores by
region"
```

```
pepper$vrsdea<-vrsdea1$eff
pepper$crsdea<-crsdea1$eff
aggregate(vrsdea~region,FUN=mean,data=pepper)
aggregate(crsdea~region,FUN=mean,data=pepper)
aggregate(vrsdea~region,FUN=min,data=pepper)
aggregate(crsdea~region,FUN=min,data=pepper)
aggregate(vrsdea~region,FUN=max,data=pepper)
aggregate(crsdea~region,FUN=max,data=pepper)
aggregate(vrsdea~region,FUN=sd,data=pepper)
aggregate(crsdea~region,FUN=sd,data=pepper)
aggregate(scale.effscore~region,FUN=mean,data=pepper)
aggregate(scale.effscore~region,FUN=min,data=pepper)
aggregate(scale.effscore~region,FUN=max,data=pepper)
aggregate(scale.effscore~region,FUN=sd,data=pepper)

"Input and output means compared to the SE using returns to scale (RTS)"
aggregate(PYIELD~SEtype1,FUN=mean,data=pepper)
aggregate(PYIELD~SEtype1,FUN=sd,data=pepper)
aggregate(PAREA~SEtype1,FUN=mean,data=pepper)
aggregate(PAREA~SEtype1,FUN=sd,data=pepper)
aggregate(PINSEC~SEtype1,FUN=mean,data=pepper)
aggregate(PINSEC~SEtype1,FUN=sd,data=pepper)
aggregate(POFERT~SEtype1,FUN=mean,data=pepper)
aggregate(POFERT~SEtype1,FUN=sd,data=pepper)
aggregate(PEAFERT~SEtype1,FUN=mean,data=pepper)
aggregate(PEAFERT~SEtype1,FUN=sd,data=pepper)
aggregate(PCFERT~SEtype1,FUN=mean,data=pepper)
aggregate(PCFERT~SEtype1,FUN=sd,data=pepper)
aggregate(PLABOR~SEtype1,FUN=mean,data=pepper)
aggregate(PLABOR~SEtype1,FUN=sd,data=pepper)

"Greenhouse pepper mutliple binary logistic regression"

"Explanatory variable selection: The exhaustive search procedure for greenhouse
peppers"
pepper$y<-pepper$grant
pepper$y<-ifelse(crsdea1$eff>=1,1,0)
library(bestglm)
res.bestglm1<-subset(pepper,select=c("rank","nutrition","otherp","rowshouse","age",
                                     "croprotation","polyethylene","gutter",
                                     "eduyears","expyears","gender",
                                     "equipmentvalue","inputsub","y"))
model1<-bestglm(Xy=res.bestglm1,family=binomial,IC="AIC",method="exhaustive")

"The choice of 5 best models"
model1$BestModels
```

```
"The selected model"
summary(model1$BestModel)

"Estimating greenhouse pepper multiple binary logistic regression"
pepper$POUT<-ifelse(crsdea1$eff>=1,1,0)
logitp<-glm(POUT~PFAMILY+PEDU+PWATER+PWELL+PEXREV+POTHER+PFARM,
            data=pepper,family="binomial")
summary(logitp)

"Logistic regression estimates in odds ratios"
library(oddsratio)
or_glm(data=data_glm,model=logitp)

"Logistic regression Nagelkerke R-squared"
library(rms)
lrm(POUT~PFAMILY+PEDU+PWATER+PWELL+PEXREV+POTHER+PFARM,data=pepper)

"Plots for the greenhouse pepper logistic regression estimates in odds ratios"
modelplot2<-plot_model(logitp,axisLabels.y=lab,transformTicks=F,sort.est=T,
                       facet.grid=F, colors=c("gray40","black"),show.values=T,
                       value.offset=0.4,digits=4, value.size=3,dot.size=1,
                       type="est",line.size=.3,
                       axis.labels=c("External revenue","Family members",
                       "Other crops grown","Irrigation in euro value",
                       "Farmer market price/kg","Education in years",
                       "Well depth in meters"),
                       title="Greenhouse Pepper Logistic Estimate Ranking",
                       axis.title="Odds Ratios",width=0.3)
modelplot2<-modelplot2+theme_set(theme_bw())
modelplot2<-modelplot2+theme(panel.grid.major=element_blank(),
                             panel.grid.minor=element_blank())
modelplot2<-modelplot2+theme(text=element_text(family="Times",size=8.5))
modelplot2

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(modelplot2)
dev.off()

"The greenhouse pepper multiple linear regression analysis"

"Explanatory variable selection: The exhaustive search procedure for greenhouse
peppers"
pepper$crsdea<-crsdea1$eff
pepper$vrsdea<-vrsdea1$eff
```

```
library(leaps)
olsmodel1<-regsubsets(crsdea~consumerprice+rank+nutrition+otherp+rowshouse+
                      irrigationvalue+croprotation+equipmentvalue+gutter+eduyears+
                      age+expyears+soiltest+grant+wholesaleprice+combinationcrops+
                      X2016usage+power+welldepth+distance,data=pepper,nbest=1,
                      nvmax=NULL,force.in=NULL,force.out=NULL,method="exhaustive")
olsmodel1
summary.out<-summary(olsmodel1)
as.data.frame(summary.out$outmat)
plot(olsmodel1,scale="adjr2")
title("Pepper Regressors based on Adjusted R-squared",adj=0.04,line=0.7,font.main=6,
      cex.main=1.4)

"The multiple linear regression model"
pepper$PDEP<-crsdea1$eff
linearp<-lm(PDEP~PFAMILY+PEDU+PWATER+PWELL+PEXREV+POTHER+PFARM,data=pepper)
summary(linearp)

"Plots for the greenhouse pepper linear regression estimates"
modelplot3<-plot_model(linearp,axisLabels.y=lab,transformTicks=F,sort.est=T,
                       facet.grid=F, colors=c("gray40","black"),show.values=T,
                       value.offset=0.4,digits=4, value.size=3,dot.size=1,
                       type="est",line.size=0.3,
                       axis.labels=c("Other crops grown","Family members",
                       "External revenue","Irrigation in euro value",
                       "Education in years","Well depth in meters",
                       "Farmer market price/kg"),
                       title="Greenhouse Pepper Linear Estimate Ranking",
                       axis.title="Estimate Values",width=0.3)
modelplot3<-modelplot3+theme_set(theme_bw())
modelplot3<-modelplot3+theme(panel.grid.major=element_blank(),
                             panel.grid.minor=element_blank())
modelplot3<-modelplot3+theme(text=element_text(family="Times",size=8.5))
modelplot3

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(modelplot3)
dev.off()

"Greenhouse pepper variable descriptive statistics for the data envelopment analysis
model"
variables1<-with(pepper,cbind(PAREA,PLABOR,PINSEC,POFERT,PEAFERT,PCFERT,PHOUSEVAL,
                 PYIELD))
library(stargazer)
stargazer(list(variables1),type="text",summary.logical=T,single.row=T,
```

```
          font.size="small",median=T,min.max=T,summary=T,digits=0)

"Coefficients of variation for the greenhouse pepper data envelopment analysis
model"
sd(pepper$PINSEC)/mean(pepper$PINSEC)
sd(pepper$PLABOR)/mean(pepper$PLABOR)
sd(pepper$PAREA)/mean(pepper$PAREA)
sd(pepper$POFERT)/mean(pepper$POFERT)
sd(pepper$PCFERT)/mean(pepper$PCFERT)
sd(pepper$PEAFERT)/mean(pepper$PEAFERT)
sd(pepper$PHOUSEVAL)/mean(pepper$PHOUSEVAL)
sd(pepper$PYIELD)/mean(pepper$PYIELD)

"Greenhouse pepper variable descriptive statistics for the linear/logistic model"
variablesloglin1<-with(pepper,cbind(PEXREV,PFARM,POTHER,PWELL,PWATER,PEDU,PFAMILY,
                       PCROP,PPOWER,PROWS,PWHOLE,PDEP,POUT))
stargazer(list(variablesloglin1),type="text",summary.logical=T,single.row=T,
          font.size="small",median=T,min.max=T,summary=T)

"Coefficients of variation for the greenhouse pepper linear/logistic model"
sd(pepper$PEXREV)/mean(pepper$PEXREV)
sd(pepper$PFARM)/mean(pepper$PFARM)
sd(pepper$POTHER)/mean(pepper$POTHER)
sd(pepper$PWELL)/mean(pepper$PWELL)
sd(pepper$PWATER)/mean(pepper$PWATER)
sd(pepper$PEDU)/mean(pepper$PEDU)
sd(pepper$PFAMILY)/mean(pepper$PFAMILY)
sd(pepper$PDEP)/mean(pepper$PDEP)
sd(pepper$POUT)/mean(pepper$POUT)
sd(pepper$PCROP)/mean(pepper$PCROP)
sd(pepper$PPOWER)/mean(pepper$PPOWER)
sd(pepper$PROWS)/mean(pepper$PROWS)
sd(pepper$PWHOLE)/mean(pepper$PWHOLE)

# Additional graphical/table representation of the results--------------------------

"Summary of the greenhouse tomato and pepper logistic regression models"
stargazer(list(logitt,logitp),type="text",summary.logical=FALSE,single.row=T,
          font.size="small")

"Summary of the greenhouse tomato and pepper linear regression models"
stargazer(list(lineart,linearp),type="text",summary.logical=FALSE,single.row=T,
font.size="small")

"Packages for the boxplot analyses"
library(ggplot2)
```

```r
library(extrafont)
loadfonts(device="win")
windowsFonts(Times=windowsFont("TTTimesNewRoman"))

"Greenhouse tomatoes: Pure technical efficiency with boxplot analysis by region"
tomato$vrsdea<-vrsdea$eff
graph<-ggplot(tomato,aes(x=region,y=vrsdea))+geom_boxplot()
graph<-graph+scale_x_discrete(name="Region")+
       scale_y_continuous(name="Pure Technical Efficiency")
graph<-graph+ggtitle("Greenhouse Tomato Efficiency Scores under VRS")
graph<-graph+theme_bw()
graph<-graph+geom_jitter()
graph<-graph+theme(plot.title=element_text(size=10))
graph<-graph+guides(fill=guide_legend(title="Region"))
graph<-graph+theme(panel.grid.major=element_blank(),
                   panel.grid.minor=element_blank())
graph<-graph+theme(text=element_text(family="Times",size=10))
graph<-graph+coord_cartesian(ylim=c(0,1))
graph<-graph+stat_summary(fun.y=mean,geom="point",shape=1,size=4)+
                          stat_summary(fun.y=mean,geom="line",aes(group=1))
graph

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(graph)
dev.off()

"Greenhouse peppers: Pure technical efficiency with boxplot analysis by region"
pepper$vrsdea<-vrsdea1$eff
graph1<-ggplot(pepper,aes(x=region,y=vrsdea))+geom_boxplot()
graph1<-graph1+scale_x_discrete(name="Region")+
        scale_y_continuous(name="Pure Technical Efficiency")
graph1<-graph1+ggtitle("Greenhouse Pepper Efficiency Scores under VRS")
graph1<-graph1+theme_bw()
graph1<-graph1+geom_jitter()
graph1<-graph1+theme(plot.title=element_text(size=10))
graph1<-graph1+guides(fill=guide_legend(title="Region"))
graph1<-graph1+theme(panel.grid.major=element_blank(),
                     panel.grid.minor=element_blank())
graph1<-graph1+theme(text=element_text(family="Times",size=10))
graph1<-graph1+coord_cartesian(ylim=c(0.2,1))
graph1<-graph1+stat_summary(fun.y=mean,geom="point",shape=1,size=4)+
                            stat_summary(fun.y=mean,geom="line",aes(group=1))
graph1
```

```
"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(graph1)
dev.off()

"Greenhouse tomatoes: Technical efficiency with boxplot analysis by region"
graph2<-ggplot(tomato,aes(x=region,y=crsdea))+geom_boxplot()
graph2<-graph2+scale_x_discrete(name="Region")+
        scale_y_continuous(name="Technical Efficiency")
graph2<-graph2+ggtitle("Greenhouse Tomato Efficiency Scores under CRS")
graph2<-graph2+theme_bw()
graph2<-graph2+geom_jitter()
graph2<-graph2+theme(plot.title=element_text(size=10))
graph2<-graph2+guides(fill=guide_legend(title="Region"))
graph2<-graph2+theme(panel.grid.major=element_blank(),
                     panel.grid.minor=element_blank())
graph2<-graph2+theme(text=element_text(family="Times",size=10))
graph2<-graph2+coord_cartesian(ylim=c(0,1))
graph2<-graph2+stat_summary(fun.y=mean,geom="point",shape=1,size=4)+
                            stat_summary(fun.y=mean,geom="line",aes(group=1))
graph2

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(graph2)
dev.off()

"Greenhouse peppers: Technical efficiency with boxplot analysis by region"
graph3<-ggplot(pepper,aes(x=region,y=crsdea))+geom_boxplot()
graph3<-graph3+scale_x_discrete(name="Region")+
                                scale_y_continuous(name="Technical Efficiency")
graph3<-graph3+ggtitle("Greenhouse Pepper Efficiency Scores under CRS")
graph3<-graph3+theme_bw()
graph3<-graph3+geom_jitter()
graph3<-graph3+theme(plot.title=element_text(size=10))
graph3<-graph3+guides(fill=guide_legend(title="Region"))
graph3<-graph3+theme(panel.grid.major=element_blank(),
                     panel.grid.minor=element_blank())
graph3<-graph3+theme(text=element_text(family="Times",size=10))
graph3<-graph3+coord_cartesian(ylim=c(0.2,1))
graph3<-graph3+stat_summary(fun.y=mean,geom="point",shape=1,size=4)+
                            stat_summary(fun.y=mean,geom="line",aes(group=1))
graph3

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
```

```
plot(graph3)
dev.off()

"Arranging together greenhouse tomato and pepper data envelopment analysis model
boxplot analyses"
require(cowplot)

"Greenhouse tomato boxplot analyses"
plot_grid(graph,graph2,align="h")

"Greenhouse pepper boxplot analyses"
plot_grid(graph1,graph3,align="h")

"Parameters for the following graphs"
par(mar=c(5.1,4.1,4.1,2.1),mgp=c(3,1,0),las=0)
par(mfrow=c(1,1))

"Greenhouse tomato BCC and CCR efficiency frontiers"
VRS<-Benchmarking::dea.plot(jitter(x1),y1,mgp=c(0.8,1,0),cex.lab=1.2,cex=1.3,
                            family="Times",xlab="Farm Inputs",ylab="Farm Output",
                            RTS="vrs",xaxt="n",yaxt="n",ORIENTATION="in-out",
                            col="Black",lty="solid",las=1,lwd=1)
CRS<-Benchmarking::dea.plot(jitter(x1),y1,mgp=c(0.8,1,0),cex.lab=1.2,cex=1.3,
                            family="Times",xlab="Farm Inputs",ylab="Farm Output",
                            RTS="crs",xaxt="n",yaxt="n", ORIENTATION="in-out",add=T,
                            col="Black",lty="longdash",lwd=1)
title("",adj=0.04,line=0.7,font.main=6,cex.main=1.2)
legend("topright",inset=0.01,box.lty=0,lty=1:2,cex=0.9,legend=c("BCC","CCR"),
       col=c("black","black"),pt.cex=1.2)
box()

"Greenhouse pepper BCC and CCR efficiency frontiers"
VRS1<-Benchmarking::dea.plot(jitter(x2),y2,mgp=c(0.8,1,0),cex.lab=1.2,cex=1.3,
                             family="Times", xlab="Farm Inputs",ylab="Farm Output",
                             RTS="vrs",xaxt="n",yaxt="n",ORIENTATION="in-out",
                             col="Black",lty="solid",las=1,lwd=1)
CRS1<-Benchmarking::dea.plot(jitter(x2),y2,mgp=c(0.8,1,0),cex.lab=1.2,cex=1.3,
                             family="Times",xlab="Farm Inputs",ylab="Farm Output",
                             RTS="crs",xaxt="n",yaxt="n",ORIENTATION="in-out",
                             add=T,col="Black",lty="longdash",lwd=1)
title("",adj=0.04,line=0.7,font.main=6,cex.main=1.2)

legend("topright",inset=0.01,box.lty=0,lty=1:2,cex=0.9,legend=c("BCC","CCR"),
       col=c("black","black"),pt.cex=1.2)
box()
```

```
# Package Citations  -----------------------------------------------------
"

'Benchmarking' Package Citation:
Bogetoft, P., & Otto, L. (2015). Benchmarking: Benchmark and Frontier Analysis Using
DEA and SFA. R package version 0.26,
https://CRAN.R-project.org/package=Benchmarking

'extrafont' Package Citation:
Chang, W. (2014). extrafont: Tools for using fonts. R package version 0.17,
https://CRAN.R-project.org/package=extrafont

'sjmisc' Package Citation:
D, Lüdecke (2018). sjmisc: Miscellaneous Data Management Tools.
R package version 2.7.1, https://CRAN.R-project.org/package=sjmisc

'sjplot' Package Citation:
D, Lüdecke (2018). sjPlot: Data Visualization for Statistics in Social Science.
R package version 2.4.1, https://CRAN.R-project.org/package=sjPlot

'rms' Package Citation:
Harrell, Frank E (2018). rms: Regression Modeling Strategies. R package version
5.1-2, https://CRAN.R-project.org/package=rms

'stargazer' Package Citation:
Hlavac, M (2015). stargazer: Well - FormattedRegression and Summary Statistics
Tables. R package version 5.2. http://CRAN.R-project.org/package=stargazer

'leaps' Package Citation:
Lumley, Th. (2017). leaps: Regression Subset Selection. R package version 3.0,
https://CRAN.R-project.org/package=leaps

'bestglm' Package Citation:
McLeod, A.I., & Xu, C. (2017). bestglm: Best subset glm using information criteria
or cross-validation. R package version 0.36,
https://CRAN.R-project.org/package=bestglm

'oddsratio' Package Citation:
Schratz, P. (2017). oddsratio: Odds ratio calculation for GAM(M)s & GLM(M)s.
R package version: 1.0.2, https://CRAN.R-project.org/package=oddsratio

'ggplot2' Package Citation:
Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York,
2009. R package version 2.2.1, https://CRAN.R-project.org/package=ggplot2

'cowplot' Package Citation:
Wilke, C. O. (2017). cowplot: Streamlined Plot Theme and Plot Annotations for
'ggplot2'. R package version 0.9.2, https://CRAN.R-project.org/package=cowplot
"
```

R PROGRAMMING SCRIPT:

ASSESSING GOVERNMENT GRANTS: EVIDENCE FROM GREENHOUSE TOMATO AND PEPPER FARMERS IN KOSOVO

```
"
    R Programming Script
    Study:  Assessing Government Grants: Evidence from Greenhouse Tomato and Pepper
    Farmers in Kosovo
    Author: Blend Frangu - Graduate student at the University of Arkansas,
            the Department of Agricultural Economics and Agribusiness.
    "
    # Reading csv data file ---------------------------------------------------

    data<-read.csv(file.choose(),header=T)

    tomato<-data[1:94, ]
    pepper<-data[95:136,]

    # Defining new variables  ------------------------------------------

    "Greenhouse tomato new variable names"
    tomato$earlygrossrevenue<-tomato$earlyprice*tomato$earlyquantity
    tomato$midgrossrevenue<-tomato$midprice*tomato$midquantity
    tomato$lategrossrevenue<-tomato$lateprice*tomato$latequantity
    tomato$seasonalrevenue<-tomato$earlygrossrevenue+tomato$midgrossrevenue+
                            tomato$lategrossrevenue

    "Greenhouse pepper new variable names"
    pepper$earlygrossrevenue<-pepper$earlyprice*pepper$earlyquantity
    pepper$midgrossrevenue<-pepper$midprice*pepper$midquantity
    pepper$lategrossrevenue<-pepper$lateprice*pepper$latequantity
    pepper$seasonalrevenue<-
    pepper$earlygrossrevenue+pepper$midgrossrevenue+pepper$lategrossrevenue

    # Results: Greenhouse tomatoes ----------------------------------------

    "Greenhouse tomato ATT based on genetic matching"
    set.seed(57)
    library(Matching)
    library(rgenoud)
    X=cbind(with(tomato,cbind(age,yield,greenhousevalue,totalarea,distance,
                              combinationcrops)))
```

```
genres<-GenMatch(Tr=tomato$grant,X=X,BalanceMatrix=X,estimand="ATT",M=3,replace=T,
                 pop.size=1000)
genATT<-Match(Y=tomato$seasonalrevenue,Tr=tomato$grant,X=X,estimand="ATT",
              Weight.matrix=genres)
summary(genATT,full=T)
balancegenATT<-MatchBalance(grant~age+yield+greenhousevalue+totalarea+distance+
                            combinationcrops, match.out=genATT,nboots=1000,
                            data=tomato)

"Estimate, lower and upper limit of the 95 percent confidence interval"
genATT$est
genATT$est-1.96*genATT$se
genATT$est+1.96*genATT$se

"ATT based on propensity score matching with logistic regression"

"Propensity score estimation model"
tomatoPS<-glm(grant~age+yield+greenhousevalue+totalarea+distance+combinationcrops,
              family=binomial(),data=tomato)
summary(tomatoPS)
round(tomatoPS$coefficients,4)
round(summary(tomatoPS)$coefficients[,2],4)

"Nagelkerke R-squared"
library(rms)
lrm(grant~age+yield+greenhousevalue+totalarea+distance+combinationcrops,data=tomato)

"Summary of the greenhouse tomato logistic regression model"
stargazer(list(tomatoPS),type="text",summary.logical=FALSE,single.row=T,
          font.size="small")

"Odds ratio"
library(oddsratio)
or_glm(data=data_glm,model=tomatoPS)

"ATT based on propensity score matching with logistic regression as the distance
measure"
modelATT<-Match(Tr=tomato$grant,Y=tomato$seasonalrevenue,X=tomatoPS$fitted,
                estimand="ATT",M=3,replace=T)
summary(modelATT,full=T)
MatchBalance(grant~age+yield+greenhousevalue+totalarea+distance+combinationcrops,
             match.out=modelATT,nboots=1000,data=tomato)

"Estimate, and lower and upper limit of the 95 percent confidence interval"
modelATT$est
modelATT$est-1.96*modelATT$se
```

```r
modelATT$est+1.96*modelATT$se

"Greenhouse tomato untreated farmers graph"
library(devtools)
library(easyGgplot2)
library(ggthemes)
library(scales)
library(ggplot2)
library(extrafont)
loadfonts(device="win")
windowsFonts(Times=windowsFont("TT Times New Roman"))
fittedvalues<-tomatoPS$fitted.values
bb<-ggplot2.histogram(data=fittedvalues,fill="white",color="black",y="density",
                      addDensityCurve=T,densityFill='gray',addMeanLine=T,
                      meanLineColor="black",binwidth=0.1,ytitleFont=c(5,"plain"))
bb<-ggplot2.customize(bb,mainTitle="Greenhouse Tomato Untreated Farmers",
                      mainTitleFont=c(12,"plain","black"),
                      xtitle="Propensity Score",xtitleFont=c(12,"plain","black"),
                      ytitleFont=c(12,"plain","black"),
                      xTickLabelFont=c(12,"plain","black"),
                      yTickLabelFont=c(12,"plain","black"))
bb<-bb+theme_bw()
bb<-bb+theme(panel.grid.major=element_blank(),
             panel.grid.minor=element_blank())
bb<-bb+theme(text=element_text(family="Times",size=10))+
       theme(axis.title.y=element_blank(),
             axis.text.y=element_blank(),
             axis.ticks.y=element_blank())
bb

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(bb)
dev.off()

"Covariate greenhouse tomato descriptive statistics"
statistics=cbind(with(tomato,cbind(age,eduyears,yield,greenhousevalue,
                      totalarea,distance,combinationcrops,seasonalrevenue)))
library(stargazer)
stargazer(list(statistics),type="text",summary.logical=T,single.row=T,
          font.size="small",median=T,min.max=T,summary=T,digits=0)

"Covariate greenhouse tomato coefficients of variation"
sd(tomato$age)/mean(tomato$age)
sd(tomato$yield)/mean(tomato$yield)
sd(tomato$greenhousevalue)/mean(tomato$greenhousevalue)
```

```
sd(tomato$totalarea)/mean(tomato$totalarea)
sd(tomato$distance)/mean(tomato$distance)
sd(tomato$combinationcrops)/mean(tomato$combinationcrops)
sd(tomato$eduyears)/mean(tomato$eduyears)
sd(tomato$seasonalrevenue)/mean(tomato$seasonalrevenue)

# Results: Greenhouse peppers -----------------------------------------

"ATT based on genetic matching"
set.seed(532)
variablesx=cbind(with(pepper,cbind(age,yield,greenhousevalue,totalarea,distance,
                                   combinationcrops)))
genres1<-GenMatch(Tr=pepper$grant,X=variablesx,BalanceMatrix=variablesx,
                  estimand="ATT",M=3,replace=T,pop.size=2000)
genATT1<-Match(Y=pepper$seasonalrevenue,Tr=pepper$grant,X=variablesx,estimand="ATT",
               Weight.matrix=genres1)
summary(genATT1)
balancegenATT1<-MatchBalance(grant~age+yield+greenhousevalue+totalarea+distance+
                             combinationcrops,match.out=genATT1,nboots=1000,
                             data=pepper)

"Estimate, and lower and upper limit of the 95 percent confidence interval"
genATT1$est
genATT1$est-1.96*genATT1$se
genATT1$est+1.96*genATT1$se

"ATT based on propensity score matching with logistic regression"

"Propensity score estimation model"
pepperPS<-glm(grant~age+yield+greenhousevalue+totalarea+distance+combinationcrops,
              family=binomial(),data=pepper)
summary(pepperPS)
round(pepperPS$coefficients,4)
round(summary(pepperPS)$coefficients[,2],4)

"Logistic regression Nagelkerke R-squared"
library(rms)
lrm(grant~age+yield+greenhousevalue+totalarea+distance+combinationcrops,data=pepper)

"Summary of the greenhouse tomato logistic regression model"
stargazer(list(pepperPS),type="text",summary.logical=FALSE,single.row=T,
          font.size="small")

"Odds ratio"
library(oddsratio)
or_glm(data=data_glm,model=pepperPS)
```

```
"ATT based on propensity score matching with logistic regression as the distance measure"
modelATT1<-Match(Tr=pepper$grant,Y=pepper$seasonalrevenue,X=pepperPS$fitted,
                 estimand="ATT",M=3,replace=T)
summary(modelATT1)
MatchBalance(grant~age+yield+greenhousevalue+totalarea+distance+combinationcrops,
             match.out=modelATT1,nboots=1000,data=pepper)

"Estimate, and lower and upper limit of the 95 percent confidence interval"
modelATT1$est
modelATT1$est-1.96*modelATT1$se
modelATT1$est+1.96*modelATT1$se

"Greenhouse pepper untreated farmers graph"
fittedvalues1<-pepperPS$fitted.values
bb1<-ggplot2.histogram(data=fittedvalues1,fill="white",color="black",y="density",
                       addDensityCurve=T,densityFill='gray',addMeanLine=T,
                       meanLineColor="black",binwidth=0.1,ytitleFont=c(5,"plain"))
bb1<-ggplot2.customize(bb1,mainTitle="Greenhouse Pepper Untreated Farmers",
                       mainTitleFont=c(12,"plain","black"),
                       xtitle="Propensity Score",
                       xtitleFont=c(12,"plain","black"),
                       ytitleFont=c(12,"plain","black"),
                       xTickLabelFont=c(12,"plain","black"),
                       yTickLabelFont=c(12,"plain","black"))
bb1<-bb1+theme_bw()
bb1<-bb1+theme(panel.grid.major=element_blank(),
               panel.grid.minor=element_blank())
bb1<-bb1+theme(text=element_text(family="Times",size=10))+
         theme(axis.title.y=element_blank(),
               axis.text.y=element_blank(),
               axis.ticks.y=element_blank())
bb1

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(bb1)
dev.off()

"Covariate greenhouse pepper descriptive statistics"
statistics1=cbind(with(pepper,cbind(age,yield,greenhousevalue,totalarea,distance,
                       combinationcrops,eduyears,seasonalrevenue)))
library(stargazer)
stargazer(list(statistics1),type="text",summary.logical=T,single.row=T,
          font.size="small",median=T,min.max=T,summary=T,digits=0)
```

```
"Covariate greenhouse pepper coefficients of variation"
sd(pepper$age)/mean(pepper$age)
sd(pepper$yield)/mean(pepper$yield)
sd(pepper$greenhousevalue)/mean(pepper$greenhousevalue)
sd(pepper$totalarea)/mean(pepper$totalarea)
sd(pepper$distance)/mean(pepper$distance)
sd(pepper$combinationcrops)/mean(pepper$combinationcrops)
sd(pepper$eduyears)/mean(pepper$eduyears)
sd(pepper$seasonalrevenue)/mean(pepper$seasonalrevenue)

# Standardized Mean Differences -------------------------------------------

"Covariate standardized mean differences using genetic matching"

library(ggalt)
library(tableone)

"Greenhouse tomatoes"

"Before treatment standardized mean differences"
xvr<-c("age","yield","greenhousevalue","totalarea","distance","combinationcrops")
beforetable<-CreateTableOne(vars=xvr,strata="grant",data=tomato,test=F)
print(beforetable,smd=T)

"After treatment standardized mean differences"
after<-tomato[unlist(genATT[c("index.treated","index.control")]),]
aftertable<-CreateTableOne(vars=xvr,strata="grant",data=after,test=F)
print(aftertable,smd=T)
tomatobeforematching<-round(c(0.017,1.378,1.544,1.038,0.412,0.681),2)
tomatoaftermatching<-round(c(0.031,0.292,0.824,0.036,0.347,0.310),2)
tomatovariables<-c("Age","Yield","Greenhouse value","Greenhouse area",
                   "Distance to the market","Other crops grown")
theme_set(theme_classic())
all<-data.frame(cbind(tomatobeforematching,tomatoaftermatching,tomatovariables))

"Greenhouse tomato covariates before matching"
library(extrafont)
loadfonts(device="win")
windowsFonts(Times=windowsFont("TT Times New Roman"))
ww<-ggplot(all,aes(tomatobeforematching,tomatovariables,
           label=paste0(tomatobeforematching),"%"))+
           geom_segment(aes(x="",y=tomatovariables,xend=tomatobeforematching,
           yend=tomatovariables),color="gray35",linetype="longdash")+
           geom_point(size=8.5,color="gray82")+
           geom_text(color="black",size=3)+xlab("Standardized Mean Difference")+
           ylab("Covariate")+ggtitle("Pre-match Greenhouse Tomato Farmers")
```

```
ww<-ww+theme(text=element_text(family="Times",size=9.7))
ww

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww)
dev.off()

"Greenhouse tomato covariates after matching"
ww1<-ggplot(all,aes(tomatoaftermatching,tomatovariables,
            label=paste0(tomatoaftermatching),
            "%"))+geom_segment(aes(x="",y=tomatovariables,xend=tomatoaftermatching,
            yend=tomatovariables),color="gray35",linetype="longdash")+
            geom_point(size=8.5,color="gray82")+geom_text(color="black",size=3)+
            xlab("Standardized Mean Difference")+ylab("Covariate")+
            ggtitle("Post-match Greenhouse Tomato Farmers")
ww1<-ww1+theme(text=element_text(family="Times",size=9.7))
ww1

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww1)
dev.off()

"Greenhouse peppers"

"Before treatment standardized mean differences"
xvr1<-c("age","yield","greenhousevalue","totalarea","distance","combinationcrops")
beforetable1<-CreateTableOne(vars=xvr1,strata="grant",data=pepper,test=F)
print(beforetable1,smd=T)

"After treatment standardized mean differences"
after1<-pepper[unlist(genATT1[c("index.treated","index.control")]),]
aftertable1<-CreateTableOne(vars=xvr1,strata="grant",data=after1,test=F)
print(aftertable1,smd=T)
tomatobeforematching1<-round(c(0.113,1.070,0.513,0.531,0.054,0.445),2)
tomatoaftermatching1<-round(c(0.759,0.534,0.637,0.728,0.287,0.136),2)
tomatovariables1<-c("Age","Yield","Greenhouse value","Greenhouse area",
                    "Distance to the market","Other crops grown")
theme_set(theme_classic())
all1<-data.frame(cbind(tomatobeforematching1,tomatoaftermatching1,tomatovariables1))

"Greenhouse pepper covariates before matching"
ww2<-ggplot(all1,aes(tomatobeforematching1,tomatovariables1,
            label=paste0(tomatobeforematching1),"%")) +
            geom_segment(aes(x="",y=tomatovariables1,xend=tomatobeforematching1,
```

```
            yend=tomatovariables1),color="gray35",linetype="longdash")+
            geom_point(size=8.5,color="gray82")+geom_text(color="black",size=3)+
            xlab("Standardized Mean Difference")+
            ylab("Covariate")+ggtitle("Pre-match Greenhouse Pepper Farmers")
ww2<-ww2+theme(text=element_text(family="Times",size=9.7))
ww2

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww2)
dev.off()

"Greenhouse pepper covariates after matching"
ww3<-ggplot(all1,aes(tomatoaftermatching1,tomatovariables1,
            label=paste0(tomatoaftermatching1),"%"))+geom_segment(aes(x="",
            y=tomatovariables1,xend=tomatoaftermatching1,yend=tomatovariables1),
            color="gray35",linetype="longdash")+geom_point(size=8.5,color="gray82")+
            geom_text(color="black",size=3)+xlab("Standardized Mean Difference")+
            ylab("Covariate")+ggtitle("Post-match Greenhouse Pepper Farmers")
ww3<-ww3+theme(text=element_text(family="Times",size=9.7))
ww3

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww3)
dev.off()

"Covariate standardized mean differences using propensity score matching with
logistic regression as the distance measure"

"Greenhouse tomatoes"

"Before treatment standardized mean differences"
xvr2<-c("age","yield","greenhousevalue","totalarea","distance","combinationcrops")
beforetable2<-CreateTableOne(vars=xvr2,strata="grant",data=tomato,test=F)
print(beforetable2,smd=T)

"After treatment standardized mean differences"
after2<-tomato[unlist(modelATT[c("index.treated","index.control")]),]
aftertable2<-CreateTableOne(vars=xvr2,strata="grant",data=after2,test=F)
print(aftertable2,smd=T)
tomatobeforematching2<-round(c(0.017,1.378,1.544,1.038,0.412,0.681),2)
tomatoaftermatching2<-round(c(0.646,0.137,0.896,0.040,0.371,0.555),2)
tomatovariables2<-c("Age", "Yield", "Greenhouse value","Greenhouse area",
                    "Distance to the market", "Other crops grown")
theme_set(theme_classic())
```

```r
all2<-data.frame(cbind(tomatobeforematching2,tomatoaftermatching2,tomatovariables2))

"Greenhouse tomato covarate standardized mean differences before matching"
ww4<-ggplot(all2,aes(tomatobeforematching2,tomatovariables2,
            label=paste0(tomatobeforematching2),"%"))+geom_segment(aes(x="",
            y=tomatovariables2,xend=tomatobeforematching2,yend=tomatovariables2),
            color="gray35",linetype="longdash")+geom_point(size=8.5,color="gray82")+
            geom_text(color="black",size=3)+xlab("Standardized Mean Difference")+
            ylab("Covariate")+ggtitle("Pre-match Greenhouse Tomato Farmers")
ww4<-ww4+theme(text=element_text(family="Times",size=9.7))
ww4

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww4)
dev.off()

"Greenhouse tomato covarate standardized mean differences after matching"
ww5<-ggplot(all2,aes(tomatoaftermatching2,tomatovariables2,
            label=paste0(tomatoaftermatching2),"%"))+geom_segment(aes(x="",
            y=tomatovariables2,xend=tomatoaftermatching2,yend=tomatovariables2),
            color="gray35",linetype="longdash")+geom_point(size=8.5,color="gray82")+
            geom_text(color="black",size=3)+xlab("Standardized Mean Difference")+
            ylab("Covariate")+ggtitle("Post-match Greenhouse Tomato Farmers")
ww5<-ww5+theme(text=element_text(family="Times",size=9.7))
ww5

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww5)
dev.off()

"Greenhouse peppers"

"Before treatment standardized mean differences"
xvr3<-c("age","yield","greenhousevalue","totalarea","distance","combinationcrops")
beforetable3<-CreateTableOne(vars=xvr3,strata="grant",data=pepper,test=F)
print(beforetable3,smd=T)

"After treatment standardized mean differences"
after3<-pepper[unlist(modelATT1[c("index.treated","index.control")]),]
aftertable3<-CreateTableOne(vars=xvr3,strata="grant",data=after3,test=F)
print(aftertable3,smd=T)
tomatobeforematching3<-round(c(0.113,1.070,0.513,0.531,0.054,0.445),2)
tomatoaftermatching3<-round(c(1.089,1.219,1.413,1.423,0.849,0.434),2)
```

```
tomatovariables3<-c("Age","Yield","Greenhouse value","Greenhouse area",
                     "Distance to the market","Other crops grown")
theme_set(theme_classic())
all3<-data.frame(cbind(tomatobeforematching3,tomatoaftermatching3,tomatovariables3))

"Greenhouse pepper covarate standardized mean differences before matching"
ww6<-ggplot(all3,aes(tomatobeforematching3,tomatovariables3,
            label=paste0(tomatobeforematching3),"%"))+
            geom_segment(aes(x="",y=tomatovariables3,xend=tomatobeforematching3,
            yend=tomatovariables3),color="gray35",linetype="longdash")+
            geom_point(size=8.5,color="gray82")+
            geom_text(color="black",size=3)+xlab("Standardized Mean Difference")+
            ylab("Covariate")+ggtitle("Pre-match Greenhouse Pepper Farmers")
ww6<-ww6+theme(text=element_text(family="Times",size=9.7))
ww6

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww6)
dev.off()

"Greenhouse pepper covarate standardized mean differences after matching"
ww7<-ggplot(all3,aes(tomatoaftermatching3,tomatovariables3,
            label=paste0(tomatoaftermatching3),"%"))+
            geom_segment(aes(x="",y=tomatovariables3,xend=tomatoaftermatching3,
            yend=tomatovariables3),color="gray35",linetype="longdash")+
            geom_point(size=8.5,color="gray82") +
            geom_text(color="black",size=3)+xlab("Standardized Mean Difference")+
            ylab("Covariate")+ggtitle("Post-match Greenhouse Pepper Farmers")
ww7<-ww7+theme(text=element_text(family="Times",size=9.7))
ww7

"High resolution graph export"
tiff(file="temp.tiff",width=3200,height=3200,units="px",res=800)
plot(ww7)
dev.off()
```

```
# Package Citations  -------------------------------------------------------
"

'ggthemes' Package Citation:
Arnold, J., Daroczi, G., Werth, B., Weitzner, B., Kunst, J., Auguie, B., Rudis, B.,
Wickham, H., Talbot, J., & London, J. (2017).
ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 3.4.0,
https://CRAN.R-project.org/package=ggthemes

'extrafont' Package Citation:
Chang, W. (2014). extrafont: Tools for using fonts. R package version 0.17,
https://CRAN.R-project.org/package=extrafont

'rms' Package Citation:
Harrell, Frank E (2018). rms: Regression Modeling Strategies. R package version
5.1-2, https://CRAN.R-project.org/package=rms

'easyGgplot2' Package Citation:
Kassambara, A. (2014). easyGgplot2: Perform and customize easily a plot with
ggplot2. R package version 1.0.0.9000, htt://www.sthda.com
'rgenoud' Package Citation:
Mebane, W. R. & Sekhon, J. S. (2017). rgenoud: R Version of GENetic Optimization
Using Derivatives. R package version 5.8-1.0,
https://CRAN.R-project.org/package=rgenoud

'ggalt' Package Citation:
Rudis, B., Bolker, B., Marwick, B., Schulz, J., Matev, R., & ProPublica. (2017).
ggalt: Extra Coordinate Systems, 'Geoms', Statistical Transformations, Scales and
Fonts for 'ggplot2'. R package version 0.4.0,
https://CRAN.R-project.org/package=ggalt

'oddsratio' Package Citation:
Schratz, P. (2017). oddsratio: Odds ratio calculation for GAM(M)s & GLM(M)s.
R package version: 1.0.2, https://CRAN.R-project.org/package=oddsratio

'Matching' Package Citation:
Sekhon, J. S. (2015). Matching: Multivariate and Propensity Score Matching with
Balance Optimization. R package version 4.9-2,
https://CRAN.R-project.org/package=Matching

'devtools' Package Citation:
Wickham, H., Hester, J., Chang, W., RStudio., & R Core team. (2018).
devtools: Tools to Make Developing R Packages Easier. R package version 1.13.5,
https://CRAN.R-project.org/package=devtools

'ggplot2' Package Citation:
Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York,
2009. R package version 2.2.1,
https://CRAN.R-project.org/package=ggplot2

'scales' Package Citation:
Wickham, H., & RStudio (2017). scales: Scale Functions for Visualization.
R package version 0.5.0, https://CRAN.R-project.org/package=scales

'tableone' Package Citation:
Yoshida, K., & Bohn, J. (2018). tableone: Create 'Table 1' to Describe Baseline
Characteristics. R package version 0.9.2,
https://CRAN.R-project.org/package=tableone
"
```