

영농교육이 귀농인의 경제적 성과에 미치는 영향: 로짓과 GBM을 활용한 성향점수매칭 분석

노은준*, 허정희**

Contents

1. 서론	51
2. 연구 방법	54
3. 연구 자료	56
4. 분석 결과	61
5. 요약 및 결론	69

Keywords

귀농인(Returning Farmers), 영농교육(Agricultural Training), 성향점수매칭(Propensity Score Matching), 경제적 성과(Economic Performance), GBM(Gradient Boosting Machine)

Abstract

본 연구는 영농교육 참여가 귀농인의 경제적 성과에 미치는 영향을 성향점수매칭(Propensity Score Matching: PSM)을 활용하여 실증적으로 분석하였다. 분석에는 2024년 창녕군과 곡성군 귀농귀촌실태조사에 응답한 귀농인 471명의 자료를 이용하였다. 영농교육 참여의 자기선택(self-selection)으로 인한 선택편의(selection bias)를 완화하기 위해 로지스틱 회귀모형과 머신러닝 기반의 Gradient Boosting Machine(GBM)을 각각 활용하여 성향점수를 추정하였으며, 성향점수매칭에는 캘리퍼(caliper)를 적용한 1:1 최근접이웃매칭 방식을 이용하였다. 성과변수는 경제생활여건 변화 정도에 대한 주관적 인식, 귀농 이후 월평균 가구소득, 귀농 전후 가구소득 변화 등 세 가지로 설정하였다. 분석 결과, 영농교육은 경제생활여건 변화에 대한 주관적 인식에는 통계적으로 유의한 영향을 미치지 않았으나, 귀농 이후 월평균 가구소득과 귀농 전후 가구소득 변화에 대해서는 분석 모형에 따라 통계적 유의성의 정도에는 차이가 있지만 일관된 정(+)의 처치효과를 나타냈다. 본 연구의 분석 결과는 영농교육이 귀농인의 소득 증대에 긍정적인 영향을 미칠 가능성을 보여주며, 따라서 정책적으로 귀농인의 영농교육 참여 기회를 확대할 필요가 있음을 시사한다.

* 경상국립대학교 대학원 농업경제학과 석사과정. e-mail: nnojoon@naver.com

** 경상국립대학교 식품자원경제학과 조교수, 농업생명과학연구원 책임연구원, 교신저자. e-mail: berliner@gnu.ac.kr

The Effects of Agricultural Training on the Economic Performance of Returning Farmers: A Propensity Score Matching Analysis Using Logit and GBM

Noh Eun-Jun*, Heo Jeonghoi**

Keywords

Returning Farmers, Agricultural Training, Propensity Score Matching, Economic Performance, Gradient Boosting Machine

Abstract

This study examines the effect of agricultural training on the economic performance of returning farmers in South Korea using Propensity Score Matching(PSM). The analysis is based on data from the 2024 Survey of Returning Farmers and Rural Migrants conducted in Changnyeong-gun and Gokseong-gun, which include 471 returning farmers. To mitigate selection bias arising from self-selection into agricultural training, propensity scores were estimated using both logistic regression(Logit) and a machine learning-based Gradient Boosting Machine(GBM). Outcome variables include subjective changes in economic living conditions, monthly household income after returning to farming, and changes in household income before and after returning to farming. The results indicate that agricultural training had no statistically significant effect on subjective evaluations of economic living conditions. However, it consistently showed positive treatment effects on average monthly household income after returning to farming and on changes in household income before and after returning to farming, despite variation in statistical significance across model specifications. These findings suggest that agricultural training positively impacts the income of returning farmers, highlighting the need for policies that expand opportunities to participate in agricultural training programs.

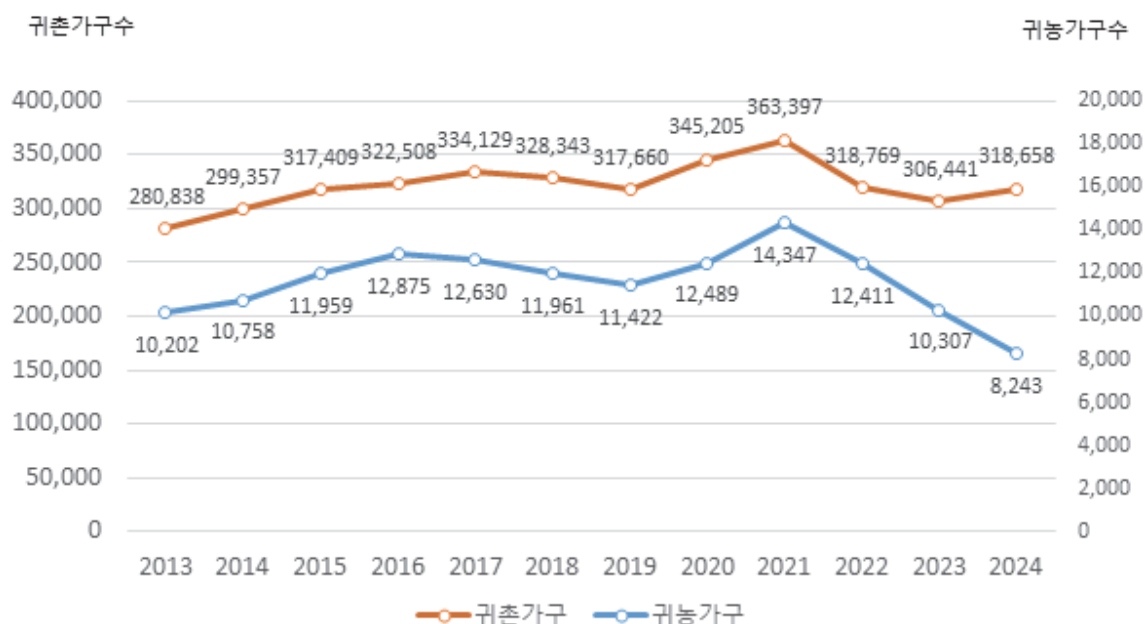
* Master's Student, Department of Agricultural Economics, Gyeongsang National University. e-mail: nnojoon@naver.com

** Assistant Professor, Department of Food and Resource Economics, Gyeongsang National University, Senior Researcher, Institute of Agriculture and Life Science, Corresponding author. e-mail: berliner@gnu.ac.kr

1. 서론

우리나라 농촌은 장기간에 걸친 인구 감소와 고령화, 청년층 유출, 지역소멸 위험의 심화로 인해 지속가능성에 대한 중대한 도전에 직면해 있다. 이러한 구조적 변화 속에서 귀농·귀촌은 단순한 주거 이전이나 개인의 삶의 방식 변화에 그치지 않고, 농촌 지역에 새로운 인구와 노동력, 소비 기반을 유입시켜 지역사회의 활력을 제고하는 중요한 정책 수단으로 인식되어 왔다. 이와 함께 귀농·귀촌 정책의 방향도 변화하고 있다. 과거에는 귀농·귀촌 인구의 양적 확대 자체가 정책의 주요 목표였다면, 최근에는 이주 이후 농촌사회에 안정적으로 정착하여 지속 가능한 생활과 소득 기반을 형성하는지를 정책 성과의 핵심 지표로 인식하고 있다. 김정섭 외(2016) 역시 귀농·귀촌 정책이 단순한 인구 유입 규모보다 정착 성과와 정책의 실효성을 중심으로 설계·평가되어야 한다고 지적한 바 있다.

그림 1. 연도별 귀농·귀촌가구 추이(2013~2024)



자료: 국가데이터처 KOSIS(2013~2024), 귀농어·귀촌인통계.

특히 귀농은 농업 생산과 농촌경제의 지속가능성에 직접적으로 기여한다는 점에서 귀촌과 구별되는 정책적 의미를 가진다. 하지만, 귀농어·귀촌인통계에 따르면 2024년 귀촌가구는 318,658가구로 전년 대비 4.0% 증가하였으나, 귀농가구는 8,243가구로 전년대비 20.0% 감소하는 등 2021년 이후 감소 추세를 보이고 있다. 이는 농촌으로의 인구 유입은 일정 수준 유지되고 있지만, 농업을 생업으로 선택하여 귀농하는 것은 여전히 높은 진입장벽이 동반되어 정착 후 안정적인 소득을 얻을 수 있는 충분한 준비와 체계적인 지원이 뒷받침될 필요가 있음을 시사한다. 2022년 귀농·귀촌 실태조사 결과에서도 귀농인의 생활 만족도와 계속 거주 의향은 비교적 높은 수준으로 나타난 반면, 귀농 초기 소득은 기존 농가의

평균에 미치지 못하는 것으로 조사되었다(농림축산식품부, 2023).

귀농 이후 안정적인 소득을 창출하기 위해서는 적합한 작목의 선택, 재배기술 습득, 경영능력 함양, 판로 확보, 정책 활용, 지역사회 적응 등 다양한 영농 역량이 요구된다. 그러나 대부분의 귀농인은 농업 경험이 부족하여 이러한 역량을 단기간에 습득하기 어려우며, 이는 초기 정착 과정에서 중요한 제약요인으로 작용한다. 이러한 배경에서 정부와 지방자치단체는 영농교육을 귀농정책의 핵심 지원수단으로 운영하고 있다. 농림축산식품부(2023)에 따르면 귀농 준비 과정에는 정착지역 탐색, 주거 및 농지 확보, 자금조달과 함께 귀농교육 참여가 주요 준비활동으로 포함되며, 귀농가구의 절반이 이주 이전에 귀농교육을 이수한 것으로 나타났다. 평균 교육시간 역시 귀농가구는 60.4시간으로 귀촌가구(48.0시간)보다 길게 나타나, 영농교육이 귀농 준비 과정에서 차지하는 비중이 상당함을 보여준다.

이처럼 귀농인의 안정적인 정착 과정에서 영농교육이 차지하는 중요성이 커지면서, 귀농교육의 역할과 성과를 규명하기 위한 연구도 다양한 측면에서 이루어져 왔다. 귀농인을 대상으로 한 기존 연구는 연구의 주요 관심을 기준으로 크게 영농교육의 역할과 효과, 귀농인의 정착 및 생활 적응, 그리고 경제적 성과와 사회·경제 활동에 관한 연구로 구분할 수 있다.

우선, 영농교육의 역할과 효과에 관한 연구이다. 마상진·김경인(2016)은 귀농교육이 농업기술과 정보 제공뿐만 아니라 인적 네트워크 형성을 통해 귀농인의 안정적인 정착에 기여한다고 분석하였으며, 교육 참여는 예비 교육시간과 자치모임 및 생산자조직 활동 정도에 의해 영향을 받는다고 보고하였다. 최윤지 외(2015)는 영농교육이 재배기술, 농업경영 능력 및 농기계 활용 능력을 향상시키는 등 귀농인의 영농역량 강화와 정착에 긍정적인 역할을 한다고 분석하였다. 또한 최원실 외(2020)는 귀농 준비교육 이수시간이 많을수록 도시로의 재이주 의향과 실제 역귀농 가능성이 낮아지는 것으로 분석하여, 영농교육이 초기 적응뿐만 아니라 장기적인 정착 유지에도 긍정적인 영향을 미칠 가능성을 제시하였다. 이러한 연구들은 영농교육이 귀농인의 영농역량을 강화하고 안정적인 정착을 지원하는 중요한 요인임을 보여준다.

다음으로, 귀농인의 정착 및 생활 적응에 관한 연구이다. 박대식·남승희(2015)는 귀농·귀촌인의 지역사회 참여 실태와 관련 요인을 분석한 결과, 이들의 지역사회 참여가 마을회의와 행사, 지역 농업인 교육, 귀농·귀촌인 모임 등을 중심으로 이루어지며, 연령, 학력, 가구소득, 준비기간, 정착연차 및 이웃과의 교류 수준 등에 따라 유의한 차이가 나타남을 밝혔다. 김경호·정성호(2018)는 귀농·귀촌인의 농촌 생활 만족도에 영향을 미치는 요인을 분석하여 소득이 중요한 사회경제적 요인으로 작용하며, 마을주민과의 관계 및 지역사회 참여 역시 생활만족도를 높이는 유의한 요인이라고 보고하였다. 박대식·김경인(2018)은 귀농·귀촌인의 직업 만족도에 영향을 미치는 요인을 분석한 결과, 소득, 건강상태, 학력, 귀농·귀촌 기간, 지역주민과의 교류 및 농업인 여부 등이 직업 만족도에 유의한 영향을 미치는 것으로 나타났다. 이러한 연구들은 귀농인의 안정적인 정착이 지역사회와의 관계 형성, 생활 적응 및 직업적 안정 등이 복합적으로 작용하는 과정임을 보여준다.

마지막으로, 귀농인의 경제적 성과와 사회·경제 활동에 주목한 연구이다. 전익수(2019)는 귀농 관련 정부정책의 소득효과를 분석한 결과, 전체 정책수혜의 평균적인 효과는 통계적으로 유의한 수준은 아니었지만 일부 주택 및 농업시설 자금지원, 체류형 지원 및 생활지원 정책은 소득에 긍정적인 영향을 미칠 수 있다고 보고하였다. 마상진·김남훈(2019)은 청년 창업농의 성공적인 정착 요인을 분석하여 초기 자본, 영농경험, 경영관리 역량, 멘토 및 지역사회와의 관계 등이 소득과 삶의 질 만족도에 중요한 영향을 미친다고 제시하였다. 한편 김정섭(2014)은 귀농인의 사회·경제 활동을 분석하면서 상당수의 귀농가구가 소규모 영농에 종사하고 있어 농업소득만으로 생계를 유지하기 어려운 경우가 많으며, 이에 따라 농외소득이나 다각적인 경제활동이 중요한 생계전략으로 기능한다고 지적하였다. 또한 귀농인의 사회·경제 활동은 개별 가구의 생계유지뿐만 아니라 농촌 지역사회의 경제적 지속가능성과 사회자본 형성에도 일정한 의미를 가진다고 보았다. 이러한 연구들은 귀농인의 성과를 생활만족이나 주관적인 적응에 한정하기보다 소득 기반의 안정성, 경제활동의 지속성 및 지역사회와의 관계 등을 포괄하는 관점에서 파악할 필요가 있음을 보여준다.

이상의 선행연구는 영농교육이 귀농인의 역량과 정착에 미치는 영향과 더불어 귀농인의 생활 적응 및 경제적 성과를 결정하는 다양한 요인을 규명하였다는 점에서 의의가 있다. 그러나 영농교육의 정책적 효과를 평가한다는 측면에서는 다음과 같은 한계가 존재한다. 먼저, 대부분의 연구는 교육 참여 실태나 교육 만족도, 교육 필요성, 교육생이 인식하는 역량 향상 등을 중심으로 분석하여 영농교육이 실제 귀농인의 경제적 성과를 얼마나 향상시키는지에 대해서는 충분히 검증하지 못하였다. 또한, 귀농성과를 분석한 연구들도 생활 만족도나 정착 의향과 같은 주관적 성과지표를 중심으로 이루어진 경우가 많아 귀농 이후 소득이나 귀농 전후 소득 변화와 같은 객관적인 경제성과를 함께 고려한 연구는 상대적으로 부족하였다.

이에 본 연구는 2024년 창녕군과 곡성군에서 실시된 귀농·귀촌 실태조사 자료를 활용하여 영농교육 참여가 귀농인의 경제적 성과에 미치는 영향을 실증적으로 분석하고자 하였다. 경제적 성과로는 경제생활여건 변화에 대한 주관적 인식과 귀농 이후 월평균 가구소득, 그리고 귀농 전후 가구소득 변화 등을 함께 고려하여 다차원적으로 접근하였다. 또한 영농교육 참여 효과를 식별하기 위해 성향점수매칭(P propensity Score Matching: PSM)을 적용하였는데, 성향점수 추정 방법으로는 기존 연구에서 널리 활용되어 온 로지스틱 회귀분석과 더불어, 변수들의 비선형성이나 복잡한 상호작용을 효과적으로 반영할 수 있는 머신러닝 기반의 Gradient Boosting Machine(GBM)을 함께 활용하였다. 이를 통해 추정 방법에 따른 분석 결과의 강건성을 검토하고, 영농교육의 정책효과에 대한 실증근거를 제시함으로써 향후 귀농교육 정책의 효과성 제고와 효율적인 정책 설계를 위한 시사점을 도출하고자 한다.

2. 연구 방법

귀농인의 영농교육 참여 여부는 무작위적으로 결정되는 것이 아니라 개인의 자발적인 선택에 의해 이루어지는 경우가 대부분이다. 따라서 교육 참여자와 비참여자는 성별, 연령, 배우자 유무, 학력, 귀농 시기, 영농형태, 영농규모뿐만 아니라 영농 의지나 귀농 준비 정도 등에서 체계적인 차이를 가질 가능성이 높다. 이러한 상황에서 두 집단간의 단순한 평균 비교나 일반적인 회귀분석만으로는 영농교육의 순수한 효과를 식별하기 어렵고, 교육 참여 여부에 내재한 선택편의(selection bias)로 인해 추정 결과가 왜곡될 수 있다.

본 연구는 이러한 문제를 완화하기 위해 Rosenbaum & Rubin(1983)이 제안한 성향점수매칭(Propensity Score Matching: PSM)을 적용하였다. 성향점수는 관측된 공변량이 주어졌을 때 특정 처치에 배정될 조건부 확률로 정의되며, 동일하거나 유사한 성향점수를 가진 처치집단과 통제집단을 비교함으로써 관측 가능한 특성의 차이를 최소화할 수 있다.¹⁾ 따라서 성향점수매칭은 무작위 실험이 어려운 정책평가 연구에서 정책 수혜집단과 비수혜집단의 비교가능성을 확보하기 위한 대표적인 준실험적(quasi-experimental) 방법으로 널리 활용되고 있다(최석준·김상신, 2009; 정해일·이상열, 2021; 김주원·박원익, 2023).

이때, 공변량 벡터 X_i 에 대해 개인 i 가 영농교육에 참여할 확률은 식 (1)과 같이 정의할 수 있다.

$$\Pr(D_i = 1 \mid X_i) \quad (1)$$

여기서 $D_i=1$ 은 영농교육에 참여한 귀농인, $D_i=0$ 은 영농교육에 참여하지 않은 귀농인을 의미한다. 한편, 성향점수매칭이 타당하게 적용되기 위해서는 조건부 독립성 가정(Conditional Independence Assumption: CIA)과 공통지지영역(Common Support Condition) 조건이 충족되어야 한다(Rosenbaum & Rubin, 1983). 조건부 독립성 가정은 $(Y_0, Y_1) \perp D \mid X$ 로 나타낼 수 있으며 관측 가능한 공변량이 주어졌을 때 처치 여부가 잠재성과 독립이라는 것을 의미한다. 공통지지영역 조건은 처치집단과 통제집단이 유사한 성향점수 구간에서 서로 비교 가능한 관측치로 구성되어야 함을 의미하며, $0 < \Pr(D_i = 1 \mid X_i) < 1$ 로 나타낼 수 있다.

성향점수 추정에는 먼저 전통적인 일반화선형모형(Generalized Linear Model: GLM) 기반의 로지스틱 회귀모형(이하 로짓모형)을 적용하였다. 로짓모형은 처치 여부가 이항형 변수인 경우 가장 널리 활용되는 성향점수 추정 방법으로, 공변량과 처치확률(p_i) 간의 관계를 식 (2)와 같은 선형으로 가정하여 영농교육 참여확률을 추정하는 방법이다.

¹⁾ 성향점수매칭은 관측 가능한 공변량을 기준으로 처치집단과 비교집단 간 선택편의를 완화하는 방법이므로 개인의 영농의지나 학습동기와 같이 관측되지 않은 요인에서 비롯되는 선택편의나 내생성 문제를 완전히 통제하기 어렵다는 방법론적 한계도 존재한다.

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (2)$$

하지만 로짓 모형은 실제 데이터에서 나타날 수 있는 비선형성이나 변수 간의 복잡한 상호작용을 충분히 반영하지 못할 가능성이 존재한다. 이에 Lee et al.(2010) 등 최근의 연구들은 비선형성과 비가법성이 존재하는 상황에서 머신러닝 방법이 로지스틱 회귀보다 더 나은 균형과 안정적인 추정 성과를 보일 수 있음을 제시하였다. 따라서 본 연구는 다양한 머신러닝 기법 가운데서도 성향점수 추정과 균형 진단에 관한 선행연구의 축적이 비교적 충분한 Gradient Boosting Machine(GBM) 모형을 선택하여, 전통적인 로지스틱 회귀모형과 함께 분석함으로써 결과의 강건성을 확보하고자 하였다.

GBM은 Friedman(2001)이 제안한 부스팅(boosting) 기반의 앙상블 학습 방법으로, 다수의 약한 학습기(weak learner)를 순차적으로 결합하여 처치확률을 추정한다. GBM은 공변량 간의 비선형 관계와 상호작용을 사전에 명시적으로 설정하지 않더라도 자료로부터 자동으로 학습할 수 있다는 장점이 있다. GBM에서 m 번째 단계의 예측함수 $f_m(x)$ 는 식 (3)과 같이 나타낼 수 있다.

$$f_m(x) = f_{m-1}(x) + v h_m(x) \quad (3)$$

이때, $f_{m-1}(x)$ 는 이전 단계까지의 누적 예측함수, $h_m(x)$ 는 현재 단계에서 적용된 약한 학습기, 그리고 v 는 학습률을 나타낸다. GBM은 높은 예측력을 갖는 반면 과적합(overfitting)이 발생할 가능성이 있으므로 모형의 복잡도를 적절히 제한할 필요가 있다. 이에 본 연구에서는 선행연구와 관련 방법론에서 권장하는 설정을 참고하여 최대 트리(tree)의 개수는 100개로 설정하였으며, 5-fold 교차검증(cross validation)을 적용하여 예측오차를 최소화하는 최적 트리 수를 선택하였다. 또한 상호작용 깊이(interaction depth)는 3, 학습률(shrinkage)은 0.005, 표본추출 비율(bag fraction)은 0.7로 설정하였다(McCaffrey et al., 2004, 2013).²⁾

이상에서 설명한 두 가지 방법을 이용해 각각 성향점수를 추정한 뒤, 이를 이용하여 처치집단과 통제집단을 매칭하였다. 본 연구에서는 추정된 성향점수를 기준으로 최근접이웃매칭(Nearest Neighbor Matching)³⁾을 실시하였으며, 처치집단의 각 관측치에 대해 성향점수가 가장 유사한 통제집단 관측치 1개를 대응시키는 1:1 비복원 매칭을 적용하였다. 또한 성향점수가 크게 상이한 관측치 간의 매칭을 방지하기 위해 캘리퍼를 적용하였으며, 추정된 성향점수 자체의 표준편차를 기준으로 캘리퍼를 각각 0.1배와 0.3배로 설정하여 분석 결과의 민감도를 검토하였다. 별도의 공통지지영역 절단은 적용하지 않았으나, 캘리퍼 범위 내에서 적절한 비교 대상을 확보하지 못한 관측치는 매칭 표본에서 제외하였다.

2) R 4.2.2의 MatchIt(4.7.2) 패키지를 통해 일반화 부스팅 모형(gbm 2.2.3)으로 추정하였다. 이때 최소 노드 크기는 옵션의 기본값(10)을 사용하였으며, 최종 선택된 트리 수는 100으로 나타났다.

3) 최근접이웃매칭의 순서는 R MatchIt 패키지의 기본 설정에 따라 성향점수가 높은 처치집단 관측치부터 순차적으로 이루어졌으며, 동일한 매칭거리를 갖는 복수의 통제집단 관측치가 존재하는 경우에는 원자료의 관측치 순서에 따라 매칭 대상을 결정하였다.

매칭 이후에는 공변량의 표준화평균차이(Standardized Mean Difference: SMD)를 이용하여 매칭 전후 로짓과 GBM 모형이 각각 어느 정도의 공변량 균형을 달성하는지를 확인하였다. 또한, SMD에 대한 보조지표로서 처치집단과 통제집단 간 공변량 분포가 전체 구간에서 얼마나 유사한지를 나타내는 경험적 누적분포함수(Empirical Cumulative Distribution Function)의 최대 차이(이하 eCDF Max)도 확인하였다.

마지막으로, 영농교육에 참여한 귀농인이 실제로 어떠한 성과를 얻었는지를 분석하기 위해 매칭된 표본을 활용하여 처치집단의 평균처치효과(Average Treatment Effect on the Treated: ATT)를 추정하였다. ATT는 영농교육에 참여한 귀농인이 교육에 참여하지 않았더라면 얻었을 것으로 예상되는 반사실적(counterfactual) 성과와 실제 성과의 차이를 평균한 개념으로 아래의 식 (4)와 같이 정의된다.

$$ATT = E[Y_1 - Y_0 \mid D = 1] \quad (4)$$

여기서 Y_1 은 영농교육에 참여했을 때의 성과, Y_0 은 영농교육에 참여하지 않았을 때의 성과를 의미한다. 본 연구에서는 ATT를 이용하여 영농교육 참여가 경제생활여건 변화 정도에 대한 주관적 인식, 귀농 이후 월평균 가구소득, 그리고 귀농 전후 가구소득 변화에 미치는 효과를 추정하였다. 평균처치효과(ATT)는 매칭가중치를 적용한 회귀모형을 통해 추정하였는데, 이때 일반적인 회귀모형의 표준오차를 적용할 경우 매칭을 통해 형성된 관측치 간 의존성을 충분히 고려하지 못할 가능성이 있으므로, 각 매칭 쌍을 하나의 군집으로 설정한 군집강건표준오차(cluster-robust standard error)를 적용하였다.

3. 연구 자료

본 연구는 경남 창녕군과 전남 곡성군에서 2024년에 실시한 ‘귀농·귀촌실태조사’ 원자료(Raw data)를 활용하였다. 귀농·귀촌실태조사는 귀농·귀촌인의 실태와 정책 수요를 파악하여 관련 정책 수립의 기초자료로 활용하기 위해 농림축산식품부가 주관하여 실시하는 국가승인통계 조사로 특별시와 광역시를 제외한 전국 9개 도 지역의 귀농인과 귀촌인 각 3,000가구를 대상으로 이루어지고 있다. 동 조사는 2016년 첫 실시 이후 2018년부터 1년 단위로 실시되고 있는데, 전국 단위 실태조사에 대한 원자료가 공개되지 않아 본 연구에서는 국가데이터처 마이크로데이터 통합서비스(MDIS)를 통해 제공되는 곡성군과 창녕군의 지자체 단위 조사 결과의 원자료를 이용하였다. 분석에 사용된 표본은 경종업에 종사하는 총 471명의 귀농인 가구주이며, 지역별로는 창녕군 243명(51.6%), 곡성군 228명(48.4%)으로 구성되어 있다. 본 조사는 귀농인의 성별, 연령, 배우자 유무, 최종학력 등 인구사회학적 특성을 비롯하여 귀농시기, 영농형태, 영농규모, 농업시작형태, 전업 및 겸업 여부, 귀농정보 획득경로 등 귀농 및 영농 관련 특성을 포함한다. 아울러 성과변수로 활용될 귀농 이전과 이후의 월평균 가구소득, 주관적

경제생활여건 변화 정도, 그리고 처치변수에 해당하는 영농교육 이수 여부에 대한 문항 등 다양한 정보를 포함하고 있다.

성향점수(Propensity Score) 추정을 위한 공변량은 인구사회학적 특성(성별, 연령, 배우자 유무, 최종 학력), 귀농 관련 특성(귀농시기, 귀농지역, 귀농정보 획득경로), 영농 관련 특성(영농형태, 영농규모, 농업시작형태, 소득구성) 등으로 구성하였다. 이러한 변수들은 귀농인의 영농교육 참여 여부(처치변수)와 귀농 이후의 경제적 성과(성과변수)에 동시에 영향을 미칠 가능성이 있는 잠재적 교란요인(Confounder)으로 판단하여 성향점수 추정 모형에 포함하였다.

공변량의 구체적인 선정 근거는 다음과 같다. 첫째, 인구사회학적 특성과 귀농 관련 특성은 귀농인의 초기 인적자본과 정보 접근성, 정착 역량을 반영하는 변수이다. 선행연구에서는 연령, 혼인상태, 학력 등이 귀농인의 정착 성과뿐만 아니라 교육 참여 가능성과도 밀접한 관련이 있는 것으로 보고되었다(허철무·하규수, 2014; 이지흠·하규수, 2017). 또한 귀농시기는 영농 경력의 차이로 이어져 이후 소득 형성에 영향을 미칠 가능성이 있으며, 귀농정보 획득경로는 교육 프로그램에 대한 접근성의 차이를 설명할 수 있다. 귀농지역(창녕군, 곡성군) 변수는 지자체별 교육 여건과 정착지원 환경의 차이를 통제하기 위해 포함하였다.

둘째, 영농 관련 특성은 귀농인의 생산 기반과 기술 수요를 반영하는 변수이다. 영농규모, 영농형태, 농업시작형태는 영농기반의 수준과 생산구조를 나타내며, 교육 참여 여부뿐 아니라 이후의 농업소득과 정착성과에도 영향을 미칠 가능성이 있다. 마상진·김남훈(2019)은 영농경력과 농지면적 등이 농업소득 및 정착 성과와 밀접한 관련이 있음을 보고한 바 있다.

한편, 본 조사에서는 영농교육 참여 시점이 귀농 이전인지 이후인지를 별도로 조사하지 않아, 일부 공변량과 교육 참여 간의 시간적 선후관계를 명확히 구분하기 어렵다는 한계가 있다. 따라서 일부 영농 관련 특성은 영농교육 참여 이후에 관측된 변수일 가능성이 있다. 성향점수매칭에서는 처치의 영향을 받을 수 있는 사후변수를 공변량으로 포함하지 않는 것이 원칙이다. 그러나 영농형태, 영농규모 및 소득구성 등은 영농교육 참여의 결과로 새롭게 결정되었다기보다 귀농 준비와 영농계획 수립 과정에서 상당 부분 사전에 형성되었을 가능성이 높은 특성으로 판단하였다. 일반적으로 귀농인은 귀농에 앞서 재배품목, 영농규모 및 영농방식 등을 계획하고, 이러한 영농 여건은 영농교육 참여 여부에도 영향을 미칠 수 있다. 이에 본 연구에서는 이들 변수가 교육 참여와 경제적 성과에 동시에 영향을 미칠 수 있는 귀농인의 영농기반 및 생산여건을 반영한다고 판단하여 성향점수 추정의 공변량에 포함하였다.

이상과 같이 본 연구에 사용된 주요 변수에 대한 기술통계량은 <표 1>에 제시하였다.

표 1. 주요 변수의 기술통계 현황

구분		단위: 명(%)		
응답자 수		창녕군	곡성군	전체
성별	남	192 (79.0)	172 (75.4)	364 (77.3)
	여	51 (21.0)	56 (24.6)	107 (22.7)
연령	19~29세	1 (0.4)	5 (2.2)	6 (1.3)
	30~39세	11 (4.5)	9 (3.9)	20 (4.2)
	40~49세	14 (5.8)	15 (6.6)	29 (6.2)
	50~59세	57 (23.5)	55 (24.1)	112 (23.8)
	60세 이상	160 (65.8)	144 (63.2)	304 (64.6)
배우자 유무	배우자 있음	208 (85.6)	187 (82.0)	395 (83.9)
	배우자 없음	35 (14.4)	41 (18.0)	76 (16.1)
최종학력	중학교 졸업 이하	53 (21.9)	50 (21.9)	103 (21.9)
	고등학교 졸업	108 (44.4)	79 (34.6)	187 (39.7)
	대학교 졸업 이상	82 (33.7)	99 (43.4)	181 (38.4)
귀농시기	2021년	100 (41.2)	109 (47.8)	209 (44.4)
	2022년	72 (29.6)	70 (30.7)	142 (30.1)
	2023년	71 (29.2)	49 (21.5)	120 (25.5)
귀농정보 획득경로 ⁴⁾	공직경로	14 (5.8)	29 (12.7)	43 (9.1)
	사적경로	63 (25.9)	76 (33.3)	139 (29.5)
	기타	166 (68.3)	123 (54.0)	289 (61.4)
영농형태	식량작물	66 (27.2)	142 (62.3)	208 (44.2)
	채소·원예	119 (49.0)	31 (13.6)	150 (31.8)
	과수	49 (20.2)	41 (18.0)	90 (19.1)
	기타	9 (3.7)	14 (6.1)	23 (4.9)
영농규모	1천 m^2 미만	3 (1.2)	24 (10.5)	27 (5.7)
	1천~3천 m^2 미만	108 (44.4)	101 (44.3)	209 (44.4)
	3천~5천 m^2 미만	51 (21.0)	46 (20.2)	97 (20.6)
	5천~10천 m^2 미만	30 (12.4)	27 (11.8)	57 (12.1)
	10천~15천 m^2 미만	27 (11.1)	18 (7.9)	45 (9.6)
	15천~20천 m^2 미만	7 (2.9)	5 (2.2)	12 (2.5)
	20천 m^2 이상	17 (7.0)	7 (3.1)	24 (5.1)
농업시작형태	가업승계형	116 (47.7)	82 (36.0)	198 (42.0)
	독립창업형	117 (48.1)	83 (36.4)	200 (42.5)
	취업·경험축적형 등 기타	10 (4.1)	63 (27.6)	73 (15.5)
소득구성	농업 수입본임	69 (28.4)	65 (28.5)	134 (28.5)
	농업 수입이 더 많음	14 (5.8)	16 (7.0)	30 (6.4)
	농업 이외 수입이 더 많음	160 (65.8)	147 (64.5)	307 (65.2)
영농교육 참여여부	참여	105 (43.2)	75 (32.9)	180 (38.2)
	미참여	138 (56.8)	153 (67.1)	291 (61.8)
경제생활 여건 변화 정도	매우 나빠짐(=1점)	5 (2.1)	6 (2.6)	11 (2.3)
	약간 나빠짐(=2점)	103 (42.4)	50 (21.9)	153 (32.5)
	보통(=3점)	116 (47.7)	132 (57.9)	248 (52.7)
	약간 좋아짐(=4점)	18 (7.4)	29 (12.7)	47 (10.0)
	매우 좋아짐(=5점)	1 (0.4)	11 (4.8)	12 (2.5)

(계속)

구분		창녕군	곡성군	전체
귀농 이후 월평균 가구소득	100만 원 미만	45 (18.5)	43 (18.9)	88 (18.7)
	100만~200만 원 미만	48 (19.8)	63 (27.6)	111 (23.6)
	200만~300만 원 미만	66 (27.2)	55 (24.1)	121 (25.7)
	300만~400만 원 미만	31 (12.8)	34 (14.9)	65 (13.8)
	400만~500만 원 미만	14 (5.8)	16 (7.0)	30 (6.4)
	500만 원 이상	39 (16.0)	17 (7.5)	56 (11.9)
귀농 전후 가구소득 변화 ⁵⁾	300만 원 이상 감소	52 (21.4)	39 (17.1)	91 (19.3)
	200만~300만 원 미만 감소	40 (16.5)	47 (20.6)	87 (18.5)
	100만~200만 원 미만 감소	48 (19.8)	48 (21.1)	96 (20.4)
	변화 없음	77 (31.7)	85 (37.3)	162 (34.4)
	100만~200만 원 미만 증가	12 (4.9)	7 (3.1)	19 (4.0)
	200만~300만 원 미만 증가	10 (4.1)	2 (0.9)	12 (2.5)
	300만 원 이상 증가	4 (1.6)	0 (0.0)	4 (0.8)

자료: 국가데이터처 MDIS(2024), 귀농·귀촌실태조사(곡성군, 창녕군).

다수의 범주형 변수를 더미변수로 모형에 포함할 경우 설명변수의 수가 지나치게 증가할 수 있으므로, 이를 완화하기 위해 일부 구간형 변수는 각 구간의 중간값을 대푯값으로 부여하여 연속형 변수로 재구성하였다. 구체적으로는 공변량 중 연령과 영농규모, 그리고 성과변수인 귀농 이후 월평균 가구소득과 귀농 전후 소득 변화에 대해 구간별 중간값을 부여하여 분석에 활용하였다.⁶⁾

표 2. 성향점수매칭 전 처치집단과 통제집단의 기초통계량

구분		Treated		Control	
		Mean	Std.Dev.	Mean	Std.Dev.
성별 (남성=1)		0.789	0.409	0.763	0.426
연령		57.458	10.806	60.649	7.505
배우자 유무(있음=1)		0.844	0.363	0.835	0.372
최종 학력	중학교 졸업 이하	0.128	0.335	0.275	0.447
	고등학교 졸업	0.389	0.489	0.402	0.491
	대학교 졸업 이상	0.483	0.501	0.323	0.468
귀농 시기	2021	0.439	0.498	0.447	0.498
	2022	0.294	0.457	0.306	0.462
	2023	0.267	0.443	0.247	0.432
귀농 지역	창녕군	0.583	0.494	0.474	0.500
	곡성군	0.417	0.494	0.526	0.500
귀농정보 획득경로	공적경로	0.128	0.335	0.069	0.253
	사적경로	0.306	0.462	0.289	0.454
	기타	0.567	0.497	0.643	0.480

4) 공적경로(군청, 귀농귀촌박람회 또는 관련 단체 및 기관), 사적경로(가족 및 지인), 기타(언론매체, 정보 없음 등).

5) 설문 문항의 귀농 이전 가구소득(범주형)과 귀농 이후 가구소득(범주형) 변수를 연속형 변수로 변환한 뒤, '귀농 이후 가구소득 - 귀농 이전 가구소득'으로 계산한 값을 활용하였다.

6) 최상위 구간의 경우 상한이 없으므로 연령은 65세, 영농규모는 22,500㎡, 가구소득은 600만 원의 값을 부여하였다. 한편, 해당 값을 달리 부여하여 민감도 분석을 실시한 결과, 추정된 ATT의 크기에는 일부 변화가 있었으나 주요 분석 결과의 방향과 통계적 유의성은 대체로 유지되었다.

(계속)

구분		Treated		Control	
		Mean	Std.Dev.	Mean	Std.Dev.
영농 형태	식량작물	0.406	0.492	0.464	0.500
	채소·원예	0.361	0.482	0.292	0.456
	과수	0.189	0.393	0.192	0.395
	기타	0.044	0.207	0.052	0.221
영농규모(m^2)		6,588.9	6,218.4	4,719.9	4,996.0
농업 시작 형태	가업승계형	0.400	0.491	0.433	0.496
	독립창업형	0.444	0.498	0.412	0.493
	기타	0.156	0.363	0.155	0.362
소득 구성	농업 수입분	0.317	0.466	0.265	0.442
	농업수입 더 많음	0.072	0.260	0.058	0.235
	농외수입 더 많음	0.611	0.489	0.677	0.468
경제생활여건 변화 정도(5점)		2.822	0.741	2.753	0.766
귀농 이후 가구소득(만 원/월)		293.056	185.641	235.052	152.818
귀농 전후 가구소득 변화(만 원/월)		-101.667	182.541	-132.302	166.209

자료: 국가데이터처 MDIS(2024), 귀농·귀촌실태조사(곡성군, 창녕군).

영농교육 참여집단(Treated)과 비참여집단(Control)의 매칭 전 특성을 비교한 결과는 <표 2>와 같다. 비교 결과, 교육 참여 여부를 결정하는 주요 공변량과 성과변수에서 두 집단 간에 유의한 차이가 존재하는 것으로 나타났다. 대표적으로 교육 참여집단의 평균 연령은 57.5세로 비참여집단 60.6세보다 약 3세 낮았으며, 최종학력은 교육 참여집단이 대졸 이상 48.3%, 중졸 이하 12.8%로 비참여집단의 대졸 이상 32.3%, 중졸 이하 27.5%에 비해 교육 수준이 더 높은 것으로 나타났다. 평균 영농규모 역시 참여집단이 6,588.9㎡, 비참여집단은 4,719.9㎡로 약 1,869㎡ 더 큰 것으로 나타났다.

또한 귀농 이후 월평균 가구소득은 참여집단이 293만 원으로 비참여집단(235만 원)보다 높았으며, 귀농 전후 가구소득 변화 폭도 참여집단(101만 원 감소)이 비참여집단(132만 원 감소)보다 작게 나타났다. 경제생활여건 변화 정도 역시 참여집단(2.82점)이 비참여집단(2.75점)보다 소폭 높았다.

이러한 매칭 전 집단 간 차이를 곧바로 영농교육의 효과로 해석하기는 어렵다. 교육 참여 여부는 개인의 연령, 영농규모, 귀농 준비 수준 등 사전 특성에 따라 선택편의가 내재될 가능성이 있으므로 이를 완화하고 영농교육의 처치효과를 보다 객관적으로 추정하기 위해 성향점수매칭(Propensity Score Matching: PSM)을 적용하였다.

4. 분석 결과

4.1. 성향점수 추정 결과

영농교육 참여에 영향을 미치는 요인을 식별하고 성향점수(Propensity Score)를 산출하기 위해 수행한 로짓모형의 추정 결과는 <표 3>과 같다. 분석 결과, 귀농인의 영농교육 참여 확률에 통계적으로 유의미한 영향을 미치는 변수는 연령, 귀농지역, 최종학력, 영농규모, 그리고 귀농정보 획득경로인 것으로 나타났다.

구체적인 변수별 추정 결과를 살펴보면, 인구사회학적 특성 중 연령의 계수 추정치는 5% 수준에서 통계적으로 유의한 부(-)의 값을 나타냈다. 이는 연령이 낮을수록 새로운 농업기술과 정보 습득을 위한 교육 참여 의지가 상대적으로 높음을 시사한다. 최종학력의 경우 대졸 이상을 기준집단으로 설정하여 분석한 결과, ‘중학교 졸업 이하’는 1% 수준에서 통계적으로 유의한 부(-)의 영향을 미치는 것으로 확인되었다. 즉, 교육 수준이 높을수록 영농교육 프로그램에 대한 접근성과 참여 가능성이 높은 경향을 보였다. 귀농지역을 나타내는 ‘창녕군’ 변수는 1% 수준에서 통계적으로 유의한 정(+)의 효과를 보였다. 이는 곡성군에 비해 창녕군 귀농인의 영농교육 참여 확률이 상대적으로 높음을 의미하며, 지역별 교육 인프라와 지원정책의 차이가 반영된 결과로 해석된다. 영농규모 역시 5% 수준에서 통계적으로 유의한 정(+)의 영향을 미쳤다. 이는 경지면적이 크고 전업적 성격이 강한 귀농인일수록 농업 생산성 향상과 경영 위험 관리에 대한 수요가 높아 영농교육에 참여할 가능성이 커짐을 시사한다.

표 3. 성향점수매칭을 위한 로짓모형 추정 결과

구분		Estimate	std.err.	z-stat.	Pr(> z)
절편		0.611	0.899	0.680	0.497
성별(기준: 여성)		0.046	0.257	0.180	0.857
연령		-0.025	0.012	-2.110	0.035**
배우자 유무 (기준: 없음)		0.111	0.294	0.380	0.704
최종학력 (기준: 대졸 이상)	중학교 졸업 이하	-0.982	0.313	-3.139	0.002***
	고등학교 졸업	-0.385	0.229	-1.683	0.092
귀농시기 (기준: 2023년)	2021년	-0.048	0.253	-0.189	0.850
	2022년	-0.026	0.272	-0.095	0.925
귀농지역 (기준: 곡성군)	창녕군	0.644	0.245	2.629	0.009***
귀농정보 획득경로 (기준: 기타)	공적경로	0.630	0.364	1.729	0.084*
	사적경로	0.165	0.245	0.672	0.501
영농형태 (기준: 기타)	식량작물	0.221	0.502	0.439	0.661
	채소·원예	0.141	0.519	0.272	0.785
	과수	0.089	0.537	0.165	0.869

(계속)

구분		Estimate	std.err.	z-stat.	Pr(> z)
영농규모		0.041	0.019	2.176	0.030**
농업시작형태 (기준: 기타)	가업승계형	-0.468	0.323	-1.451	0.147
	독립창업형	-0.198	0.323	-0.612	0.540
소득구성 (기준: 농외소득 더 많음)	농업수입뿐	0.219	0.239	0.918	0.358
	농업수입 더 많음	0.406	0.422	0.962	0.336

Log-likelihood = -289.8322, LR χ^2 = 46.8733, (Prob) χ^2 = 0.000221, Pseudo R² = 0.0748, No. of obs. = 471

주: ***, **, *는 각각 1%, 5%, 10%에서 통계적으로 유의함을 나타냄.
자료: 저자 작성.

GBM 모형 추정 결과 도출된 공변량의 상대적 중요도(Relative Importance)는 <표 4>와 같다. 변수 중요도는 로짓 모형의 회귀계수처럼 개별 변수의 영향 방향(+, -)이나 선형적 한계효과를 직접적으로 나타내지는 않지만, 성향점수 추정 과정에서 어떤 변수가 처치 확률(영농교육 참여) 예측에 상대적으로 크게 기여했는지를 직관적으로 보여준다.

<표 4>의 분석 결과를 살펴보면, GBM 모형에서 가장 높은 중요도를 기록한 변수는 ‘영농규모’였으며, 그 뒤를 이어 ‘연령’, ‘중학교 졸업 이하’, ‘창녕군’ 순으로 나타났다. 특히 영농규모와 연령의 상대적 중요도 합은 60.8%에 달해, 머신러닝 기반의 GBM 모형이 영농교육 참여 확률을 예측할 때 이 두 변수를 가장 핵심적인 판단 기준으로 활용했음을 보여준다. 여기에 중학교 졸업 이하(학력 특성)와 창녕군(지역 특성)까지 포함할 경우, 상위 4개 핵심 변수의 누적 중요도는 88.9%에 이른다. 이는 앞서 수행한 로짓 모형에서 통계적으로 유의미하게 나타났던 핵심 변수들이 GBM 모형에서도 여전히 비중있게 예측력을 발휘하고 있음을 보여준다.

반면, 이 외의 공변량들은 모형에 미치는 영향력이 상대적으로 제한적이었다. 독립창업형(농업시작 형태), 식량작물, 채소·원예(이상 영농형태), 공적경로(귀농정보 획득경로) 등은 처치 예측에 기여하는 정도가 상대적으로 낮았고, 나머지 변수들 역시 미미한 중요도를 나타냈다.

표 4. GBM 모형의 변수 중요도 분석 결과

순위	변수명	중요도(%)	순위	변수명	중요도(%)
1	영농규모	34.64	10	사적경로	0.63
2	연령	26.20	11	가업승계형	0.59
3	중학교 졸업 이하	17.50	12	2022년	0.47
4	창녕군	10.57	13	농업 수입분임	0.41
5	독립창업형	2.22	14	성별	0.38
6	식량작물	2.10	15	농업 수입이 더 많음	0.36
7	공적경로	1.60	16	배우자 유무	0.19
8	채소·원예	1.14	17	과수	0.19
9	고등학교 졸업	0.83	18	2021년	0.00

자료: 저자 작성.

4.2. 균형성 검정(Balancing Test) 결과

성향점수매칭(P propensity Score Matching) 이후, 영농교육 참여집단(처치집단)과 비참여집단(통제집단) 간의 공변량 분포가 유의미하게 동질해졌는지를 확인하기 위해 균형성 검정(Balancing Test)을 수행하였다. 본 연구에서는 매칭 전후 두 집단 간 공변량의 동질성 정도를 통계적으로 진단하기 위해 표준화평균차이(Standardized Mean Difference, 이하 SMD)를 활용하였다(표 5)). SMD는 두 집단 간 평균 또는 비율의 차이를 표준편차로 나누어 산출한 값으로, 통상적으로 해당 지표의 절댓값이 0.1 미만이면 공변량의 균형 상태가 양호한 것으로 판단한다(Austin, 2009).

표 5. 성향점수매칭 전후 공변량 균형 비교: 표준화평균차이

변 수		매칭 전	매칭 후(Logit)		매칭 후(GBM)	
			caliper = 0.1	caliper = 0.3	caliper = 0.1	caliper = 0.3
성향점수		0.6160	0.0185	0.0810	0.0154	0.0512
성별 (남성=1)		0.0637	-0.1257	-0.1167	-0.0160	0.0151
연령		-0.2953	-0.0569	-0.1256	0.0889	0.0131
배우자 유무 (있음=1)		0.0259	0.0884	0.0493	0.0541	0.0170
최종학력	중학교 졸업 이하	-0.4407	-0.0576	-0.0535	0.0979	0.0000
	고등학교 졸업	-0.0270	0.0131	0.0488	-0.0804	0.0000
귀농시기	2021	-0.0158	-0.0775	-0.0720	-0.0922	-0.0871
	2022	-0.0250	-0.0141	0.0000	0.0143	0.0677
귀농지역	창녕군	0.2213	-0.0780	-0.0483	0.0530	0.0000
귀농정보 획득경로	공적경로	0.1769	-0.0384	0.0000	-0.0196	0.0925
	사적경로	0.0367	0.1113	0.0905	0.0426	0.0804
영농형태	식량작물	-0.1189	0.0000	0.0000	0.0399	0.0000
	채소·원예	0.1437	-0.0133	-0.0372	0.0544	0.0900
	과수	-0.0091	-0.0328	0.0000	-0.0668	-0.0631
영농규모		0.3006	-0.0253	0.0311	-0.0888	-0.0690
농업시작 형태	가업승계형	-0.0673	-0.0654	-0.0486	-0.0800	-0.0756
	독립창업형	0.0645	0.0129	0.0000	0.0526	0.0745
소득구성	농업 수입뿐임	0.1119	-0.0827	-0.0512	-0.0422	0.0000
	농업 수입이 더 많음	0.0533	0.0495	0.0690	0.1010	0.0954

자료: 저자 작성.

아울러 본 연구에서는 허용 반경(Caliper) 설정에 따른 매칭 결과의 민감도를 검토하고 실증 분석의 강건성(Robustness)을 확보하고자, 캘리퍼를 성향점수 추정치의 표준편차를 기준으로 0.1배와 0.3배 두 가지로 설정하였다. 일반적으로 캘리퍼 기준이 넓을수록 더 많은 표본을 매칭에 활용할 수 있다는 장점이 있는 반면, 캘리퍼 기준을 좁고 엄격하게 설정할수록 공변량의 균형을 더욱 정교하게 확보할 수 있다. 이처럼 서로 다른 캘리퍼 기준을 적용함으로써 추정 결과가 특정 매칭 조건에 과도하게 의존하는지

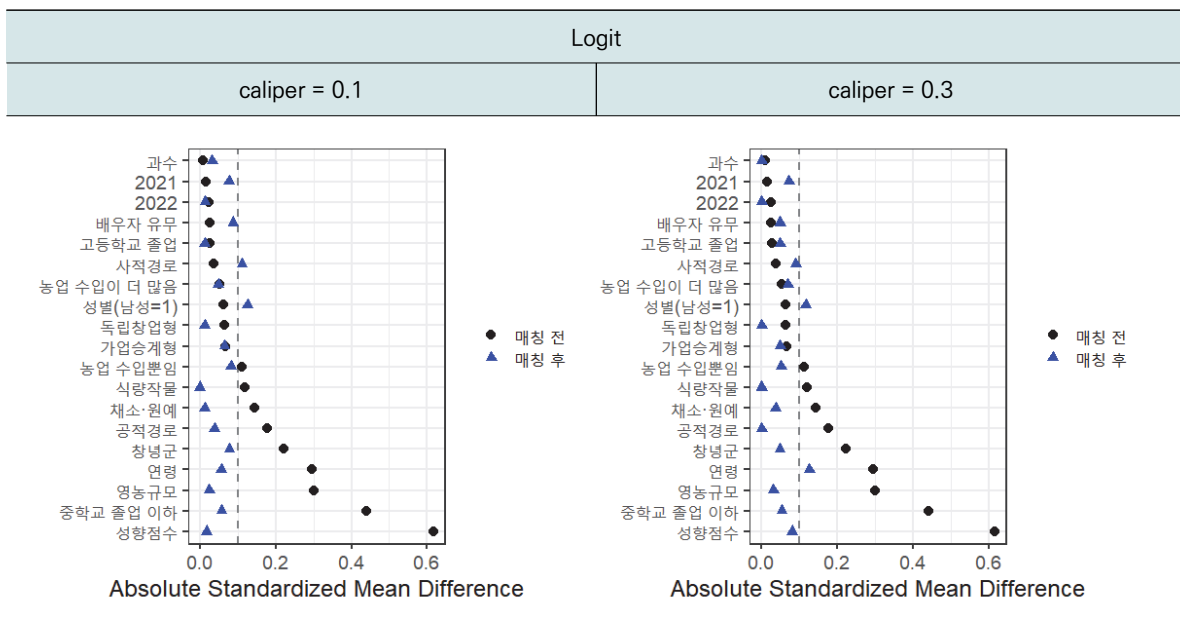
점검하고, 표본 유지와 균형 개선 간의 절충 관계를 종합적으로 검토하여 분석의 신뢰성을 높일 수 있도록 하였다.

균형성 검정 결과, 두 모형 모두 매칭 이후 성향점수의 표준화평균차이(SMD)가 크게 감소하였음을 확인할 수 있다. 특히 캘리퍼를 0.3에서 0.1로 엄격하게 조정했을 때, 로짓과 GBM 두 모형 모두 성향점수의 분포 격차가 추가로 축소되었다. 성향점수 자체의 SMD는 로짓 모형의 경우 0.0810에서 0.0185로, GBM 모형은 0.0512에서 0.0154로 낮아져, 엄격한 캘리퍼 설정이 전반적인 균형도 제고에 효과적이었음을 보여준다.

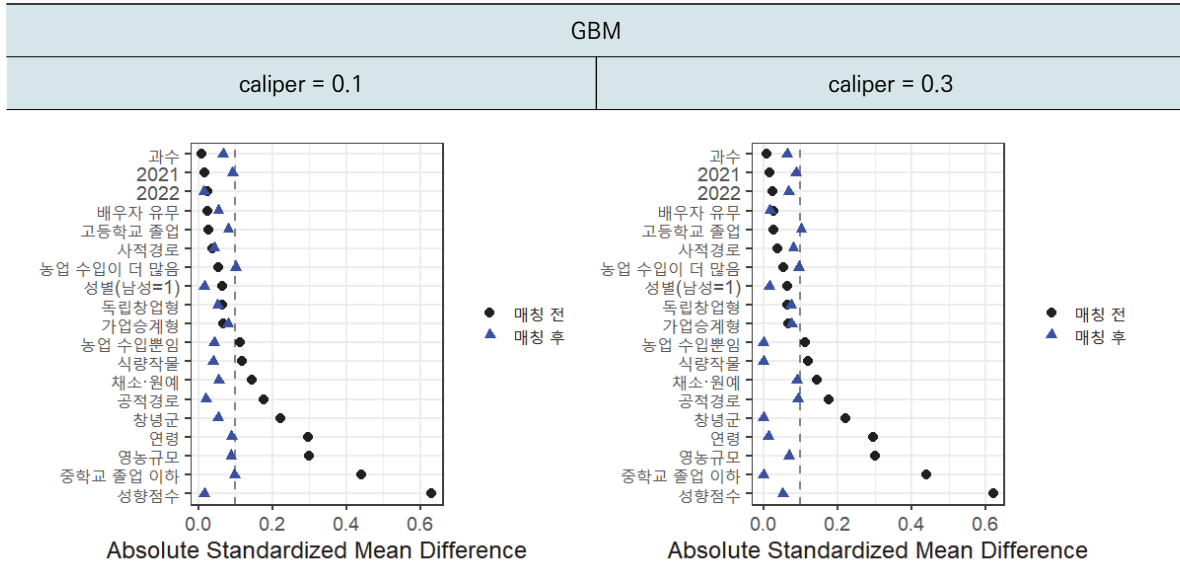
개별 공변량 수준에서도 이러한 경향성이 일관되게 관측되었다. 0.3 캘리퍼 기준에서도 두 모형 모두 전반적으로 양호한 균형 상태를 달성하였으나, 0.1 캘리퍼를 적용하였을 때 연령, 성별, 영농규모 등 주요 변수에서의 잠재적 불균형이 한층 더 완화된 양상을 보였다. 다만 모든 변수에서 0.1 캘리퍼가 절대적으로 우세한 것은 아니며, 일부 범주형 변수에서는 오히려 0.3 캘리퍼 조건에서 더 안정적인 추정치를 보이기도 하였다. 즉, 엄격한 매칭 기준이 전반적인 균형 개선에는 유리하지만, 세부 변수별 조정 성과는 모형의 특성과 공변량의 데이터 형태에 따라 다소 상이할 수 있음을 시사한다.

〈그림 2〉는 성향점수매칭 전후 공변량의 표준화평균차이(SMD)를 시각화한 그래프(Love plot)이다. 매칭 전에는 영농규모, 연령, 최종학력(중학교 졸업 이하), 지역(창녕군) 등 일부 변수에서 표준화평균차이가 특히 크게 나타나 두 집단 간 불균형이 존재하였으나 매칭 이후에는 SMD가 전반적으로 0에 가까워지면서 균형이 개선되었음을 확인할 수 있다.

그림 2. 매칭 전후 표준화평균차이 Love plot



(계속)



자료: 저자 작성.

표준화평균차이와 더불어 공변량의 균형성을 평가하는 데 활용되는 경험적 누적분포함수 (empirical cumulative distribution function)의 최대 차이(이하 eCDF Max)는 처치집단과 통제 집단 간 공변량의 분포가 전체 구간에서 얼마나 유사한지를 나타내는 지표이다. 두 집단 간 평균의 차이를 중심으로 평가하는 SMD와 달리, eCDF Max는 공변량의 분포 전반에서 나타나는 차이를 파악할 수 있다는 점에서 균형성 평가의 보조지표로 활용되며, eCDF Max 값이 0에 가까울수록 두 집단의 공변량 분포가 유사함을 의미한다. 본 연구에서는 공변량 균형성 평가에서 일반적으로 활용되는 SMD의 0.1 기준을 준용하여 eCDF Max 역시 0.1 이하를 양호한 균형의 판단 기준으로 활용하였다. <표 6>에 제시된 바와 같이 매칭 전에는 성향점수, 영농규모, 중학교 졸업 이하, 연령, 지역(창녕군) 등의 변수에서 eCDF Max 값이 상대적으로 크게 나타나 처치집단과 통제집단 간 분포에 차이가 존재하는 것으로 나타났다. 반면, 매칭 이후에는 대부분 공변량의 eCDF Max 값이 뚜렷하게 감소하였으며, 대체로 0.1 이하 수준으로 낮아져 두 집단 간 공변량의 분포 균형이 전반적으로 개선되었음을 확인할 수 있다.

표 6. 성향점수매칭 전후 공변량 균형 비교: eCDF Max

변 수		매칭 전	매칭 후(Logit)		매칭 후(GBM)	
			caliper = 0.1	caliper = 0.3	caliper = 0.1	caliper = 0.3
성향점수		0.2887	0.0385	0.0893	0.0458	0.0926
성별 (남성=1)		0.0260	0.0513	0.0476	0.0065	0.0062
연령		0.1257	0.0641	0.0893	0.0915	0.0370
배우자 유무 (있음=1)		0.0094	0.0321	0.0179	0.0196	0.0062
최종 학력	중학교 졸업 이하	0.1471	0.0192	0.0179	0.0327	0.0000
	고등학교 졸업	0.0132	0.0064	0.0238	0.0392	0.0494

(계속)

변 수		매칭 전	매칭 후(Logit)		매칭 후(GBM)	
			caliper = 0.1	caliper = 0.3	caliper = 0.1	caliper = 0.3
귀농 시기	2021	0.0078	0.0385	0.0357	0.0458	0.0432
	2022	0.0114	0.0064	0.0000	0.0065	0.0309
귀농지역	창녕군	0.1091	0.0385	0.0238	0.0261	0.0000
귀농 정보 획득 경로	공직경로	0.0590	0.0128	0.0000	0.0065	0.0309
	사적경로	0.0169	0.0513	0.0417	0.0196	0.0370
영농 형태	식량작물	0.0584	0.0000	0.0000	0.0196	0.0000
	채소·원예	0.0690	0.0064	0.0179	0.0261	0.0432
	과수	0.0036	0.0128	0.0000	0.0261	0.0247
영농규모		0.1995	0.0641	0.0893	0.0719	0.0679
농업 시작 형태	가업승계형	0.0330	0.0321	0.0238	0.0392	0.0370
	독립창업형	0.0321	0.0064	0.0000	0.0261	0.0370
소득 구성	농업 수입뿐임	0.0521	0.0385	0.0238	0.0196	0.0000
	농업 수입이 더 많음	0.0138	0.0128	0.0179	0.0261	0.0247

자료: 저자 작성.

한편, 분석에 투입되는 표본의 유지 측면에서는 예상대로 0.3 캘리퍼가 상대적으로 유리한 것으로 나타났다. 매칭의 기준이 되는 캘리퍼를 0.3에서 0.1로 축소함에 따라, 최종 분석에 활용된 매칭 표본은 로짓 모형에서 168쌍에서 156쌍으로 감소하였고, GBM 모형은 162쌍에서 153쌍으로 줄어들었다. 이는 0.1 캘리퍼 적용 시 공변량 균형은 보다 정교하게 개선되지만, 엄격한 반경 조건으로 인해 공통지지영역(Common Support)을 벗어나 제외되는 표본 손실(Sample Loss)이 증가하기 때문이다.

표 7. 매칭 표본 수

구분	Logit				GBM			
	caliper = 0.1		caliper = 0.3		caliper = 0.1		caliper = 0.3	
	Control	Treated	Control	Treated	Control	Treated	Control	Treated
ALL	291	180	291	180	291	180	291	180
Matched	156	156	168	168	153	153	162	162
Unmatched	135	24	123	12	138	27	129	18

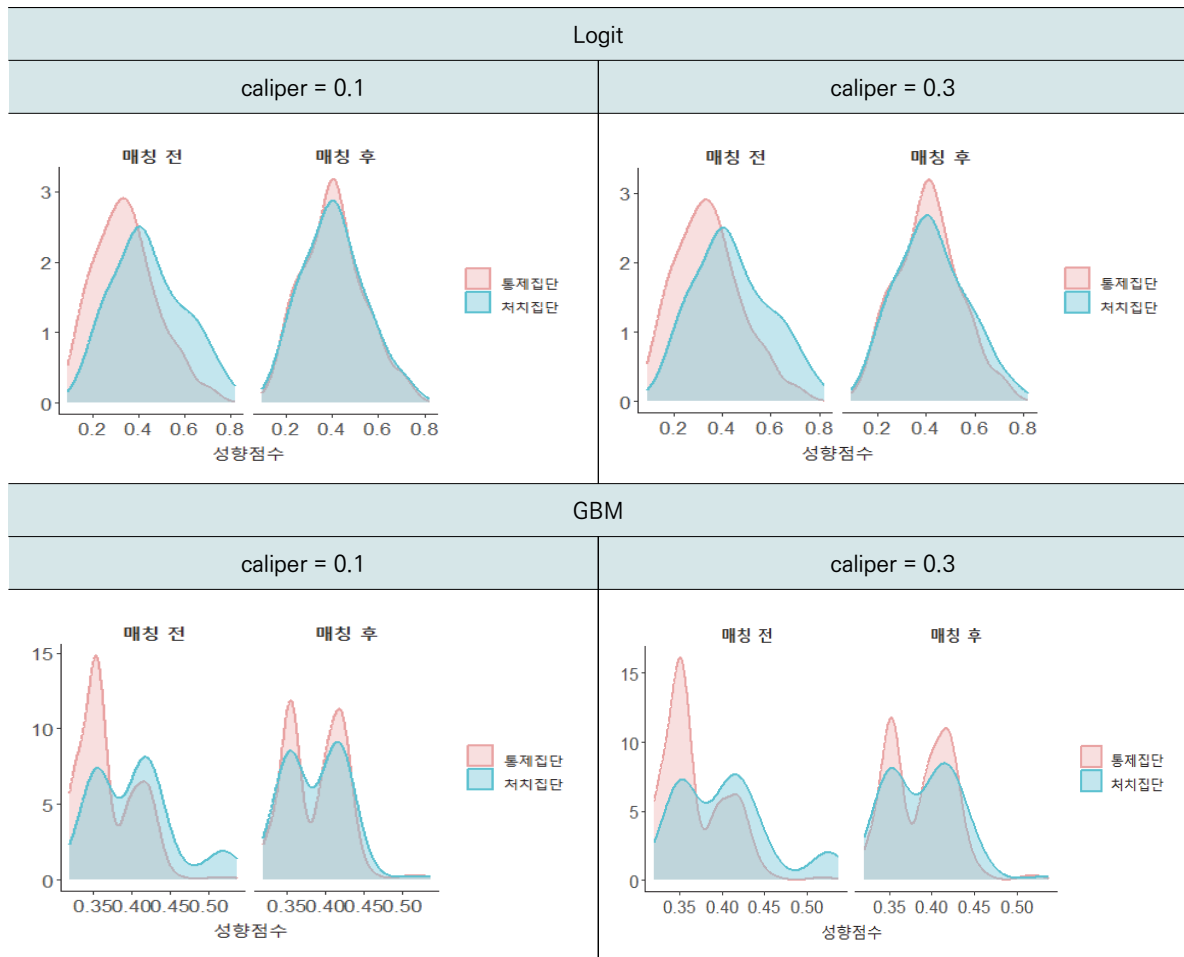
자료: 저자 작성.

〈그림 3〉은 성향점수매칭(PSM) 전후 처치집단과 통제집단의 성향점수 분포 변화를 비교함으로써, 공변량 균형성 개선 양상을 시각적으로 제시하는 확률밀도함수 그래프(Density Plot)이다. 먼저 로짓 모형 추정 결과를 살펴보면, 매칭 전에는 처치집단과 통제집단의 확률밀도분포가 서로 다른 위치에 치우쳐 있고 분포 형태에서도 뚜렷한 차이를 보여 두 집단 간 공변량 분포의 불균형과 선택편의의 가능성이 존재함을 시사한다. 그러나 매칭 이후에는 두 집단의 밀도곡선이 전반적으로 유사한 형태를 보이며, 특히 중심 구간에서의 중첩 정도가 현저하게 증가하였다. 이는 성향점수매칭을 통해 공통지지영역 내

에서 비교가능한 표본이 효과적으로 구성되었으며, 두 집단 간 공변량 균형이 크게 개선되었음을 보여 준다. 허용 반경별로는 캘리퍼 0.1과 0.3 모두 뚜렷한 균형 개선 효과를 보였으나, 보다 엄격한 기준을 적용한 0.1 조건에서 분포의 정렬이 한층 더 정교하게 일치되는 양상을 보였다.

GBM 추정 결과는 선형 기반의 로짓 모형과는 상이한 분포 양상을 보인다. 매칭 전에는 처치집단과 통제집단의 성향점수 분포가 다봉형(multimodal) 구조를 나타내며 복수의 국소적 밀집구간에서 차이를 보였다. 그러나 매칭 이후에는 다봉형 분포의 특성이 유지되는 가운데 두 집단의 밀도곡선이 대부분의 구간에서 높은 중첩을 나타내었으며, 성향점수도 상대적으로 좁은 구간에 집중되는 양상을 보였다.

그림 3. 매칭 전후의 성향점수분포



자료: 저자 작성.

4.3. ATT 추정 결과

이상의 균형성 검정을 통해 매칭 이후 처치집단과 통제집단 간 공변량 균형이 전반적으로 개선된 것으로 판단됨에 따라, 매칭된 표본을 활용하여 영농교육이 귀농인의 경제적 성과에 미친 평균처치효과(Average Treatment Effect on the Treated: ATT)를 추정하고 그 결과를 <표 8>에 제시하였다. ATT는 1:1 비복원 매칭표본을 이용하여 추정하였으며, 매칭된 관측치에는 동일한 가중치를 적용하였다. 표준오차는 각 처치-통제 매칭쌍을 하나의 군집으로 설정하여 산출한 군집강건표준오차(cluster-robust standard errors)를 사용하였다.⁷⁾

표 8. 영농교육의 평균처치효과(ATT) 추정 결과

구 분		경제생활여건 변화 정도 (5점 척도)		귀농 이후 월평균 가구소득 (만 원/월)		귀농 전후 가구소득 변화 (만 원/월)	
		Logit	GBM	Logit	GBM	Logit	GBM
caliper = 0.1	ATT	0.013 (0.082)	0.078 (0.084)	33.333* (18.957)	47.386*** (17.506)	33.974* (20.379)	38.562* (21.618)
	95% CI	-0.149, 0.174	-0.086, 0.243	-3.822, 70.489	13.074, 81.697	-5.968, 73.917	-3.808, 80.932
caliper = 0.3	ATT	0.006 (0.084)	0.111 (0.078)	37.500** (17.906)	45.370** (17.822)	41.369** (18.698)	36.111* (20.368)
	95% CI	-0.159, 0.170	-0.042, 0.264	2.405, 72.595	10.439, 80.302	4.721, 78.017	-3.808, 76.031

주: ***, **, *는 각각 1%, 5%, 10%에서 통계적으로 유의함을 의미하며, 괄호 안의 수치는 군집강건 표준오차를 나타냄.

자료: 저자 작성.

우선 주관적 체감 지표인 ‘경제생활여건 변화 정도(5점 척도)’에 대한 분석 결과를 살펴보면, 영농교육의 평균처치효과(ATT)는 모든 분석모형에서 매우 작은 수준으로 추정되었으며 통계적으로도 유의하지 않은 것으로 나타났다. 구체적으로 영농교육 참여자의 경제생활여건 변화 정도 점수는 비참여자에 비해 최대 약 0.1점 높은 정도에 그쳤다. 5점 척도 응답임을 고려하면 이러한 차이는 실질적인 효과로 해석하기 어려운 수준이며, 통계적 유의성 또한 확인되지 않았다. 따라서 영농교육이 귀농 초기 정착 과정에서 경제생활여건에 대한 주관적 평가나 전반적인 체감 여건을 단기적으로 개선하는 효과는 제한적인 것으로 판단된다.

하지만 객관적 경제 성과 지표인 ‘귀농 이후 월평균 가구소득’에서는 모든 분석 모형에서 일관되게 정(+)의 효과가 추정되었다. ATT는 칼리퍼 설정값에 따라 로짓 모형에서 33.33만~37.50만 원, GBM에서 45.37만~47.39만 원으로 나타났다. 이는 영농교육 비이수자 그룹의 귀농 이후 월평균 가구소득이 235만 원 수준임을 고려할 때, 영농교육 이수자 그룹의 월평균 가구소득이 약 14.2~20.2% 높은 수

7) ATT는 R의 lm 함수를 이용하여 추정하였으며, 군집강건 표준오차는 sandwich 패키지의 vcovCL 함수를 이용하여 산출하였다.

준으로 추정되었음을 의미한다. 특히 GBM 캘리퍼 0.1에서는 1% 유의수준에서 유의성이 확인되었고, 상대적으로 많은 표본을 유지한 0.3 캘리퍼에서는 로짓과 GBM 모형 모두 5% 유의수준에서 통계적 유의성이 확인되어, 영농교육이 귀농인의 소득 향상에 긍정적인 영향을 미쳤을 가능성을 뒷받침한다.

이러한 경제적 성과 개선 경향은 또 다른 소득 지표인 '귀농 전후 가구소득 변화'에서도 일관되게 나타났다. 해당 변수의 ATT는 로짓 모형에서 33.97만~41.37만 원, GBM에서 36.11만~38.56만 원으로 추정되었다. 귀농·귀촌실태조사에 따르면 통상적으로 귀농 이후에는 귀농 전에 비해 가구소득이 감소하는 것으로 나타났는데, 분석 결과에 따르면 영농교육 참여 시 비참여 집단에 비해 이러한 가구소득 감소 폭이 약 34만~41만 원 줄어드는 효과가 있었음을 의미한다. 이상의 결과를 종합하면, 영농교육은 경제생활여건 변화에 대한 주관적 인식에서는 통계적으로 유의한 효과가 확인되지 않았으나, 귀농 이후 월평균 가구소득과 귀농 전후 가구소득 변화 등 객관적인 경제성과 지표에서는 일부 통계적 유의성의 차이는 있지만 일관되게 긍정적인 효과를 보인 것으로 나타났다. 이러한 결과는 영농교육을 통해 습득한 재배기술과 농업경영 역량, 그리고 지역 내 농업 네트워크의 형성이 생산성과 소득 향상으로 이어졌을 가능성을 시사한다.

5. 요약 및 결론

본 연구는 2024년 경남 창녕군과 전남 곡성군에서 실시된 귀농·귀촌 실태조사 자료와 성향점수매칭(P propensity Score Matching: PSM) 방법을 활용하여 영농교육 참여가 귀농인의 경제적 성과에 미치는 효과를 실증적으로 분석하였다. 이를 위해 로짓(Logit) 모형과 머신러닝 기반 Gradient Boosting Machine(GBM) 두 가지 방법을 이용하여 성향점수매칭을 실시하고, 이를 토대로 영농교육 참여 여부에 따른 그룹 간의 평균처치효과(ATT)를 추정하였다.

주요 분석 결과는 다음과 같다. 먼저, 귀농인의 영농교육 참여는 연령, 학력, 영농규모, 지역 특성 등의 사전적 특성과 밀접한 관련을 갖는 것으로 나타났다. 로짓 모형뿐만 아니라 GBM에서도 이러한 변수들의 상대적 중요도가 높게 나타나 교육 참여가 귀농인이 보유한 자원과 정착 여건에 의해 영향받고 있음을 확인할 수 있었다.

영농교육의 평균처치효과(ATT) 추정 결과, 주관적 경제생활여건 변화에는 통계적으로 유의한 영향을 미치지 않았지만, 객관적인 경제적 성과에서는 일관된 긍정적 효과를 나타냈다. '경제생활여건 변화 정도'는 모든 분석 시나리오에서 그룹 간의 유의한 차이가 확인되지 않았는데, 이는 귀농인의 주관적 만족도가 교육 참여 여부뿐 아니라 지역사회 적응, 생활환경, 사회적 관계 등 다양한 요인의 영향을 받기 때문으로 해석된다. 반면 '귀농 이후 월평균 가구소득'에서는 모든 분석모형에서 정(+)의 처치효과가 확인되었으며 영농교육 참여자가 비참여자에 비해 약 33만~47만 원이 높아 그룹 간의 평균소득 수준을

기준으로 볼 때 약 14.2~20.2% 정도의 소득 상승 효과가 나타나는 것으로 추정되었다. 또한, '귀농 전후 가구소득 변화'에 대한 ATT 추정 결과는 통상적으로 귀농 이후 소득이 귀농 이전보다 감소하는 경향을 고려할 때, 영농교육 참여 집단의 귀농 이후 월평균 가구소득 감소 폭이 비참여 집단에 비해 약 34만~41만 원 작은 것으로 해석할 수 있다. 이러한 결과는 영농교육이 영농기술 습득, 농업경영 역량 강화, 인적 네트워크 형성 등을 통해 귀농인의 생산성과 소득 향상에 기여하였을 가능성을 시사한다.

본 연구의 분석 결과 귀농인의 영농교육 참여가 소득 성과에 유의한 영향을 미치는 것으로 나타난 만큼, 향후 정책적으로는 귀농인의 영농교육 참여 기회를 확대하고 교육 참여 제약 요인을 완화하는 데 초점을 둘 필요가 있는 것으로 판단된다. 본 연구의 전체 응답자 471명 가운데 영농교육 참여자는 38.2%에 그치고 있다. 영농교육 참여가 소득 성과의 향상으로 이어질 수 있다는 분석 결과를 고려하면, 정책의 우선적인 과제는 교육의 효과가 기대되는 귀농인들이 실제 교육에 접근하고 참여할 수 있도록 여건을 조성하고 참여를 유도하는 데 있다고 볼 수 있다.

아울러 조사 대상자들의 인구사회학적 특성을 살펴보면, 60세 이상이 64.6%로 가장 높은 비중을 차지하고 있으며, 영농규모 역시 1천 m² 이상 3천 m² 미만이 44.4%로 가장 큰 비중을 차지하고 있다. 이처럼 고령층과 소규모 영농인을 중심으로 한 귀농 구조가 나타나고 있는 만큼, 영농교육의 확대 역시 일률적인 교육 공급보다는 귀농인의 특성을 고려하는 방향으로 이루어질 필요가 있다. 구체적으로, 고령 귀농인이 부담 없이 참여할 수 있도록 교육 시간과 방식을 탄력적으로 운영하고, 소규모 영농인의 영농 여건과 수요를 반영한 접근성 높은 프로그램을 제공하는 한편, 초기 정착 기반이 취약한 귀농인에 대해서는 교육 관련 정보 제공과 참여 유도를 강화할 필요가 있다.

본 연구는 다음과 같은 한계도 지닌다. 첫째, 원자료 확보의 제약으로 인해 분석 대상이 2개의 군 단위 지자체에 한정되었으며, 이로 인해 연구 결과를 전국 단위로 일반화하는 데에는 신중한 해석이 요구된다. 둘째, 영농교육의 경제적 성과를 보다 정확하게 분석하기 위해서는 농업소득을 성과지표로 활용하는 것이 바람직하나, 본 연구에서 활용한 귀농·귀촌 실태조사에서는 농업소득을 별도의 조사항목으로 포함하고 있지 않아 부득이 가구소득을 대리변수로 활용하였다. 셋째, 귀농인이 이수한 영농교육은 교육 시간이나 내용 등에 따라 양적, 질적으로 차이가 있을 수 있으나, 자료의 제약으로 인해 이러한 교육의 이질성을 분석에 충분히 반영하지 못하였다. 마지막으로, 영농교육 참여 시점이 귀농 이전인지 이후인지 구분하여 조사하지 않아, 일부 공변량과 교육 참여 간의 시간적 선후관계를 명확히 확인하기 어렵다는 한계가 있다. 그럼에도 불구하고 해당 공변량은 그 특성상 귀농 및 영농 준비 과정에서 상당 부분 사전에 결정되었을 가능성이 높다고 판단하여 분석에 포함하였다.

본 연구는 영농교육이 귀농인의 객관적인 경제적 성과, 특히 소득 향상에 긍정적인 영향을 미칠 수 있음을 실증적으로 제시하였다는 점에서 의미가 있다. 이는 최근 귀농정책이 단순한 귀농 인구의 확대에서 안정적인 정착과 경제적 자립을 지원하는 방향으로 전환되고 있는 가운데, 이러한 정책 기조를 뒷받침하는 실증적 근거를 제시한다는 점에서 학술적, 정책적 의의를 가진다. 향후에는 전국 단위의 패널자

료를 활용하여 영농교육 효과의 지속성과 지역별 이질성을 보다 면밀하게 분석할 필요가 있으며, 이를 추후 연구과제로 남겨둔다.



이 저작물은 크리에이티브 커먼즈 저작자표시-비영리-변경금지 4.0 국제 라이선스에 따라 이용할 수 있습니다.

This work is licensed under CC BY-NC-ND 4.0.

참고문헌

- 국가데이터처 KOSIS. (2013~2024). 귀농어·귀촌인통계. <<https://kosis.kr>>.
- 국가데이터처 MDIS. (2024). 곡성군귀농·귀촌실태조사. <<https://mdis.mods.go.kr/index.do>>
- 국가데이터처 MDIS. (2024). 창녕군귀농·귀촌실태조사. <<https://mdis.mods.go.kr/index.do>>
- 김경호, 정성호. (2018). 귀농·귀촌인의 농촌 생활만족도에 영향을 미치는 요인. *지역사회학*, 19(1), 5-26.
- 김정섭, 마상진, 김종인, 오정훈. (2016). *귀농·귀촌 종합계획 수립 방향 연구*. 한국농촌경제연구원.
- 김정섭. (2014). 귀농인의 사회·경제활동과 함의. *농촌지도와 개발*, 21(3), 53-89.
<https://doi.org/10.12653/jecd.2014.21.3.0053>
- 김주원, 박원익. (2023). 성향점수매칭(PSM)-이중차분(DID) 결합모형을 이용한 혁신조달 정책의 기업지원 효과 분석. *기술혁신연구*, 31(3), 201-230. <https://doi.org/10.14386/sime.2023.31.3.201>
- 농림축산식품부. (2023. 3. 2.). *2022년 귀농귀촌 실태조사 결과 발표*. <<https://www.mafra.go.kr/bbs/home/792/565547/artclView.do>>.
- 마상진, 김경인. (2016). 귀농·귀촌인의 교육 참여 영향 요인 탐색. *농업교육과 인적자원개발*, 48(2), 83-104.
<http://doi.org/10.23840/agehrd.2016.48.2.83>
- 마상진, 김남훈. (2019). 청년 창업농의 성공적 정착에 미치는 요인 분석. *농촌경제*, 42(1), 27-49.
<http://doi.org/10.36464/jrd.2019.42.1.002>
- 박대식, 김경인. (2018). 귀농·귀촌인의 직업 만족도 영향 요인 분석. *농업교육과 인적자원개발*, 50(3), 41-60.
<http://doi.org/10.23840/agehrd.2018.50.3.41>
- 박대식, 남승희. (2015). 귀농·귀촌인의 지역사회참여 실태와 관련 요인. *농촌사회*, 25(1), 41-87.
 UCI: G704-000099.2015.25.1.003
- 이지흠, 하규수. (2017). 귀농 귀촌에서 귀농만족에 미치는 영향요인 분석: 가족 지지도를 조절효과 중심으로. *벤처창업연구*, 12(5), 39-53. <http://doi.org/10.16972/apjbve.12.5.201710.39>
- 전익수. (2019). 정부 정책이 귀농인 가구소득에 미친 효과 분석. *농촌경제*, 42(1), 109-135.
<http://doi.org/10.36464/jrd.2019.42.1.005>
- 정해일, 이상열. (2021). 성향점수매칭(PSM)과 이중차분법(DID)을 활용한 정부 지원 경영컨설팅 사업의 효과 분석. *회계와 정책연구*, 26(2), 237-260. <http://doi.org/10.21737/RAPS.2021.05.26.2.237>
- 최석준, 김상신. (2009). 성향점수 매칭을 이용한 정부 연구개발 보조금 효과분석. *한국과학기술학회논문지*, 10(1), 200-208. UCI: G704-001653.2009.10.1.002
- 최원실, 허태호, 이상현. (2020). 역 귀농·귀촌 의향과 결정에 미치는 요인 분석. *농촌경제*, 43(3), 65-90.
<http://doi.org/10.36464/jrd.2020.43.3.004>
- 최윤지, 한송희, 공민재. (2015). 귀농 교육의 만족도, 성취도, 효과성 분석에 관한 연구. *농촌계획*, 21(1), 51-61.
<http://doi.org/10.7851/ksrp2015.21.1.051>
- 허철무, 하규수. (2014). 귀농자의 귀농만족도에 미치는 영향 요인 연구. *벤처창업연구*, 9(2), 233-250.
<http://doi.org/10.16972/apjbve.9.2.201404.233>
- Austin, P. C. (2009). Balance Diagnostics for Comparing the Distribution of Baseline Covariates Between Treatment Groups in Propensity-Score Matched Samples. *Statistics in Medicine*, 28(25), 3083-3107.
<https://doi.org/10.1002/sim.3697>
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R. & Burgette, L. F. (2013). A Tutorial on Propensity Score Estimation for Multiple Treatments Using Generalized Boosted Models. *Statistics in Medicine*, 32(19), 3388-3414. <https://doi.org/10.1002/sim.5753>
- McCaffrey, D. F., Ridgeway, G. & Morral, A. R. (2004). Propensity Score Estimation With Boosted Regression

- on for Evaluating Causal Effects in Observational Studies. *Psychological Methods*, 9(4), 403-425. <https://doi.org/10.1037/1082-989X.9.4.403>
- Rosenbaum, P. R. & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41-55. <https://doi.org/10.1093/biomet/70.1.41>
- Lee, K. L., Lessler, J. & Stuart, E. A. (2010). Improving Propensity Score Weighting Using Machine Learning. *Statistics in Medicine*, 29(3), 337-346. <https://doi.org/10.1002/sim.3782>