

## RESEARCH ARTICLE

# Personalized digital extension services and agricultural performance: Evidence from smallholder farmers in India

Pallavi Rajkhowa<sup>1\*</sup>, Matin Qaim<sup>1,2</sup>

**1** Center for Development Research (ZEF), University of Bonn, Bonn, Germany, **2** Institute for Food and Resource Economics, University of Bonn, Bonn, Germany

\* [pallavirajkhowa@uni-bonn.de](mailto:pallavirajkhowa@uni-bonn.de)

## Abstract

Productivity growth in smallholder agriculture is an important driver of rural economic development and poverty reduction. However, smallholder farmers often have limited access to information, which can be a serious constraint for increasing productivity. One potential mechanism to reduce information constraints is the public agricultural extension service, but its effectiveness has often been low in the past. Digital technologies could enhance the effectiveness of extension by reducing outreach costs and helping to better tailor the information to farmers' individual needs and conditions. Using primary data from India, this study analyses the association between digital extension services and smallholder agricultural performance. The digital extension services that some of the farmers use provide personalized information on the types of crops to grow, the types and quantities of inputs to use, and other methods of cultivation. Problems of selection bias in the impact evaluation are reduced through propensity score matching (PSM) combined with estimates of farmers' willingness to pay for digital extension. Results show that use of personalized digital extension services is positively and significantly associated with input intensity, production diversity, crop productivity, and crop income.

## OPEN ACCESS

**Citation:** Rajkhowa P, Qaim M (2021) Personalized digital extension services and agricultural performance: Evidence from smallholder farmers in India. *PLoS ONE* 16(10): e0259319. <https://doi.org/10.1371/journal.pone.0259319>

**Editor:** Bjorn Van Campenhout, International Food Policy Research Institute, UNITED STATES

**Received:** July 16, 2021

**Accepted:** October 16, 2021

**Published:** October 28, 2021

**Copyright:** © 2021 Rajkhowa, Qaim. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting Information](#) files.

**Funding:** Pallavi Rajkhowa: Financial support by the Deutscher Akademischer Austauschdienst (DAAD) and the Federal Ministry for Economic Cooperation and Development (BMZ) under the Center for Development Research (ZEF) Project "Measures to reduce price volatility for improved food security" is gratefully acknowledged. The funders had no role in study design, data collection

## 1. Introduction

In developing countries, productivity growth in small-farm agriculture can serve as an important driver of economic development and poverty reduction [1,2]. However, smallholder farmers typically face many challenges, such as unpredictable weather conditions, market risks, and limited access to information, technologies, and financial services. These and other constraints result in low productivity and low rates of market participation [3]. Hence, a key policy question for promoting rural development and poverty reduction is how the main information and market access constraints that smallholder farmers face can be overcome.

In most developing countries, agricultural extension services are the dominant method of public-sector support towards knowledge diffusion and innovation in the small-farm sector [4]. Traditionally, extension agents have either tried to educate farmers directly about best practices or have worked with selected "model farmers" who are then expected to act as

and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

information multipliers [5]. However, the effectiveness of traditional extension approaches has been limited, either because of too little funding and thus low outreach or information that is not sufficiently tailored to farmers' needs [4,5]. The development and use of new digital extension approaches, which build on information and communication technologies (ICTs) such as mobile phones and internet platforms, could potentially improve the situation, but empirical evidence of actual impacts is scarce. In this article, we use data from smallholder farmers in India to analyse whether digital extension with personalized advice can help to increase innovation, productivity, and income.

There is a growing body of literature highlighting that digital technologies can lower the cost of communication significantly, thus reducing transaction costs, increasing market efficiency, and promoting economic growth and poverty reduction [6–11]. One strand of the literature focuses particularly on the role of mobile phones in agricultural development. For instance, several studies showed that the use of mobile phones can increase selling and reduce buying prices [12–14], enhance smallholder market access [15–18], and increase farm productivity and income [19–21]. Some studies also analysed the impacts of mobile phones on other dimensions of smallholder welfare, showing positive effects on nutrition and gender equality [22,23], off-farm employment [24], and migration [25]. Most of these studies focus on mobile phones as a simple communication tool. A few other studies looked at the effects of mobile phone-based financial services, such as mobile money, on-farm performance, and household welfare [26,27]. Overall, these studies suggest that mobile phones can be very beneficial for smallholder farmers.

Over the last few years—with the rise of high-speed internet connections and web-enabled smartphones—the use of ICTs has further evolved. Recent studies analysed the use of the internet and smartphones in rural areas of developing countries, generally finding positive effects on household welfare [28–30]. Various internet-based applications and technologies are being developed, which could have major implications for agricultural development. Cloud services, low-cost open-source software, and big data analytics contribute to the emergence of new, internet-based *agricultural technology platforms* (agri-tech platforms henceforth) that aggregate supply and demand by reducing the number of intermediaries and connecting farmers directly to agro-advisory services, input providers, retailers, or consumers. Rao et al. [31] reviewed such agri-tech platforms in India and categorized them into those that connect farmers to (i) extension and agro-advisory services, (ii) input suppliers, and (iii) buyers of agricultural produce. Similar platforms are also emerging in many other developing countries.

While various types of ICTs are increasingly used in agricultural extension, the literature on how these new digital extension services affect smallholder performance remains relatively thin. Several studies showed that providing farmers with general market and weather information through mobile phones, text messages, or internet applications can promote agricultural productivity and market efficiency [6,20,21,32–35]. There are also a few studies that analysed the effects of using training videos or call centres and interactive voice response services for farmers, with somewhat mixed results [19,36,37]. However, in these examples, ICTs were used primarily to improve the delivery of generic extension information. We are aware of only two published studies that analysed the effects of personalized advice (as opposed to standardized generic advice), and both these studies were carried out with farmers in Nigeria [38,39]: both studies showed that site-specific nutrient management recommendations provided through digital decision support tools help to increase crop yields. We contribute to this literature by analysing the association between the use of personalized digital extension services and smallholder agricultural performance in India. Unlike Arouna et al. [38] and Oyinbo et al. [39], who focused on nutrient management advice only, the digital application that we study involves a broader extension package, including advice for vegetable farmers on what species

and varieties to grow, what quantities of different chemicals to use, and how to obtain these inputs at good quality and prices.

Digital technologies can help to better tailor different components of agricultural advice to farmers' individual needs in multiple ways. For instance, predictive analytics and machine learning algorithms can be used to combine data on weather forecasts, soil conditions, market prices, and other aspects to develop and deliver site-specific agricultural recommendations. Theoretically, such digital extension services can affect smallholder households through several mechanisms. First, they can reduce information barriers by providing personalized advice on which types of crops to grow in what season, the appropriate types and quantities of inputs to use, and the best timing for the different operations and input applications. Second, they can connect farmers to new input markets by providing transparent information on local market prices and reputed brands and suppliers. Third, they can help improve farmers' bargaining power by providing transparency and additional supplier options. Fourth, improved access to personalized information and new technologies and inputs can increase the levels of commercialization. These mechanisms will likely change farmers' cropping patterns and increase their input intensities, crop yields, sales volumes, and incomes.

In this study, we use the example of a concrete digital extension application in India to analyse whether a positive association between personalized advice and farm performance can be observed. For the study, a survey of smallholder vegetable farmers was conducted in early-2019. Some of the farmers surveyed had already adopted the digital extension services, while others had not. Using propensity score matching and a variety of other statistical techniques, our results confirm positive and significant associations between the adoption of personalized digital advice and various farm performance indicators, including input intensity, production diversity, vegetable productivity, and vegetable income.

The rest of this article is organized as follows. Section 2 describes the survey region in India and concrete features of the digital extension services, followed by an explanation of the sampling strategy and the outcome variables. Section 3 discusses the econometric strategy for the impact evaluation. Section 4 presents and discusses the results, while Section 5 concludes.

## 2. Materials and methods

### 2.1 Ethics statement

This study was approved by the Research Ethics Committee of the Center for Development Research (ZEF), University of Bonn. Verbal consent was obtained from each study participant.

### 2.2 Study region and digital extension services

We focus on one large Farmer Producer Organisation (FPO) named 'Mayurbhanj Agri Smart Farmer Producer Company Limited' in Mayurbhanj District in the state of Odisha, eastern India. The FPO comprises around 1000 farmers growing vegetables, rice, and a few other crops. It was initiated in 2017 by the National Bank for Agriculture and Rural Development (NABARD), a public financial institution in India, and eKutir, an India-based Public Benefit Corporation. eKutir develops and establishes digital platforms to create more efficient markets for smallholder farmers in India and other parts of Asia and Africa. eKutir also developed the digital agri-tech platform 'Farmex' (<https://farm-ex.io/>), which is of particular interest here. At the time of our data collection in 2019, the agri-tech platform was called 'Bloom'. The name was changed to 'Farmex' in 2021.

Farmex offers vegetable farmers real-time agricultural extension services and a marketplace for vegetable seeds, fertilizers, and pesticides. In the future, the digital services shall be extended to the output market as well, but when we collected the data in early-2019 this was

not yet the case. As part of the extension services, Farmex helps its users to plan season-wise cropping activities and provides information on best practices for growing specific vegetable crops. The platform also offers recommendations on the types and quantities of inputs to use and on relevant pests and diseases and how to control them. Moreover, to reduce issues with the use of counterfeit inputs, which are widespread in India, the platform makes suggestions on specific input brands and suppliers.

Due to low levels of education and widespread digital illiteracy among the FPO members, farmers do not receive the extension services on their mobile phones. Instead, when farmers want to adopt the digital extension services, they get in touch with the head of the FPO who then operates the internet-based application on the farmer's behalf. In other words, the FPO head takes on the role of an extension agent, equipped with the digital technology that enables him/her to provide tailor-made agricultural advice and services to the FPO members.

Adoption of the digital extension services is voluntary for farmers and is currently free of charge. In this context, adoption is defined as the farmers' active decision to subscribe to the agri-tech platform with all available services included. If a member of the FPO decides to adopt the services, the FPO head creates an individual account by entering personalized data, including farm-specific details such as location, land size, types of crops currently grown, and soil conditions. These details—together with the application's algorithms on weather forecasts, market conditions, and optimal production decisions—are processed to provide personalized advice on crop selection, the schedule of agricultural activities, and input regimes. After every season, the FPO head enters additional data on the actual inputs used by each farmer, the yields obtained, and the prices to further improve the algorithms' predictions and advice for future seasons. Data entry and updating are typically done during personal meetings between the FPO head and individual farmers. However, once the data are entered into the system, much of the actual advice is given through phone calls and text messages, thus lowering transaction costs.

### 2.3 Sampling strategy

Data for this study were collected through a survey conducted between January and March 2019. The selected FPO in Odisha has members in two blocks (Betnoti and Badasahi) and 26 villages. Out of all 26 villages, we randomly selected 20 villages (10 in each of the two blocks) for data collection (the remaining six villages were used for pre-testing the questionnaire). In each of the sampled villages, a household census was conducted, and all households were categorized into three groups (i) FPO vegetable farmers, (ii) non-FPO vegetable farmers, and (iii) non-FPO non-vegetable farmers. For this study, we were only interested in vegetable farmers because the digital extension services are related to vegetable cultivation. Hence, we only selected farmers from the first two groups on a random basis. The total sample includes 1105 vegetable-growing households, out of which 603 were members of the FPO and 502 were not. This distribution is proportional to the actual population proportions (S1 Table in the supporting information).

The digital extension services are accessible only to FPO members. However, as adoption for FPO members is voluntary, not all FPO members subscribed to the digital extension services. Of the 603 FPO members in our sample, around 77% (465) had adopted digital extension services by early-2019, the others had not although they would have been eligible. To analyse the association between the use of digital extension services and agricultural performance, we only include those farmers that had adopted digital extension services as part of the "treatment" group. Hence, the control group includes 640 farm households that did not adopt digital extension services irrespective of their FPO membership. It is important to mention

that—beyond the digital agri-tech platform—none of the farmers in the sample villages had access to other types of formal extension services.

Relevant questions with our sampling strategy are why many of the vegetable farmers in the target villages were not members of the FPO and whether it is appropriate to include these non-members in the control group. In this context, it is important to stress that the FPO in Odisha is open to all farmers in the respective villages who grow vegetables, meaning that there are no eligibility criteria in terms of farm size or other factors. Farmers decide themselves whether they want to become FPO member. From the existing literature on farmer organisations, we know that farmers' membership decisions typically depend on expected costs and benefits, which are influenced by various socioeconomic characteristics [40]. However, as explained above, the FPO in Odisha was only initiated in 2017 (less than two years prior to the survey), so that many farmers were still considering to join. [S2 Table](#) in the supplementary information compares various socioeconomic characteristics between FPO members and non-members in the control group, showing that both types of farms are similar. Hence, pooling FPO members and non-members in the control group seems justified. The existing farmer heterogeneity is addressed with econometric techniques, as is explained further below.

## 2.4 Data collection

Data from each randomly selected farm household were collected through personal interviews with the person responsible for farm management (mostly the household head) using a structured questionnaire. The interviews were conducted in the local language (Oriya) by trained enumerators who were supervised by the researchers. Before the actual survey, the questionnaire was pretested and adjusted by interviewing 60 households in the six non-sampled villages in the FPO area.

All agricultural data were collected for the 12 months from March 2018 to February 2019 to capture all seasons of the year. Details on crop production were asked separately for the Kharif, Rabi, and Zaid seasons. These seasonal data were summed up later to calculate annual input and output variables. As is well known, recall data collected from smallholder farmers on land sizes and yield are prone to measurement error, which can lead to bias in the statistical analysis [41]. In our context, measurement error is possible but of lesser concern than in many other smallholder studies, because we focus primarily on vegetable production, a commercial enterprise for farmers in the study region, for which written farm records and receipts are typically available. In addition to the agricultural data, information on various household characteristics, other economic activities, perceptions about digital technologies, and social networks was gathered. These data are used to control for possible confounding factors in the analysis.

## 2.5 Outcome variables

We want to analyse the association between using digital extension services and agricultural performance. Agricultural performance is evaluated in terms of crop production diversity, input use intensity, crop productivity, crop commercialization, and crop income. All outcomes are measured over 12 months covering the Kharif, Rabi, and Zaid seasons. The outcome variables are defined more specifically as:

- i. *Crop production diversity*: Farmers in the study area traditionally grow rice and vegetables. One of the stated objectives of the digital agri-tech platform and extension services is to help farmers diversify their production by growing more types of vegetables. We measure crop production diversity by counting the number of different crop species grown on the farm during the 12 months. The effect of the digital extension services on production diversity

could be positive if farmers learn about growing new types of vegetables, but it could also be negative if they specialize in particularly lucrative species.

- ii. *Input use intensity*: Input use intensity is measured in terms of the monetary expenditures for seeds, fertilizer, pesticides, and all inputs combined per acre of cropland. Input use intensity is of interest because the digital extension services provide specific advice on the types and quantity of inputs to use. Effects could be positive if farmers previously under-invested in inputs, but they could also be negative if farmers previously overused certain inputs or paid too high prices due to information asymmetry.
- iii. *Crop productivity*: Productivity is measured in terms of the monetary value of the output produced per acre of land, whereby the total land cultivated by the farm is considered. We use monetary values because farmers grow many different types of crops for which physical weights are not easily comparable. Improved access to information through digital extension services is expected to result in higher crop productivity. Negative effects on productivity are not expected, as personalized advice on crop choices and input regimes will hardly lead to lower yields and revenues, unless the information provided is flawed.
- iv. *Crop commercialization*: Crop commercialization is defined here as the share of total crop output sold during the 12 months. Many farmers keep some of their output for home consumption. We use average market prices for the particular crops produced to value home-consumed quantities. As the digital extension services are expected to increase productivity and also provide better access to market information, the effect on crop commercialization is also expected to be positive. As the extension provided only refers to crops, we focus on crop commercialization alone and do not include the livestock sector.
- v. *Crop income*: Annual crop income is calculated as the gross value of crop production (including the output not sold valued at average market prices) minus variable production costs (purchased seeds, fertilizer, pesticides, manure, hired labour, irrigation water, machinery, and transportation). It is expected that higher productivity and higher levels of commercialization will also lead to higher crop income. Farmers would hardly adopt and use digital extension services if these were associated with lower crop income.

### 3. Econometric approach

#### 3.1 Modelling the adoption of digital extension services

Let the decision to adopt digital extension services be a dichotomous choice, such that  $D_i = 1$  if household  $i$  adopts the digital extension services and  $D_i = 0$  otherwise. Smallholders choose to adopt when the expected utility from using the services ( $U_{iD}$ ) is greater than the utility from not using them ( $U_{iN}$ ), such that  $U_{iD} > U_{iN}$ . The difference between the utility achieved from adopting and not adopting can be denoted by a latent variable  $Z^*$ , such that  $Z^* = [(U_{iD}) - (U_{iN})] > 0$ . Since  $Z^*$  is a latent variable, it is unobservable [42]. However, it can be expressed in terms of observed variables as follows:

$$Z_i^* = \beta X_i + \varepsilon_i, \quad D_i = 1 [Z_i^* > 0] \quad (1)$$

where  $\beta$  is a vector of parameters to be estimated,  $X_i$  is a vector of household, farm, and contextual characteristics, and  $\varepsilon_i$  is an error term that is assumed to be normally distributed.

The probability of households adopting the digital extension services can be expressed as:

$$\Pr(D_i = 1 | X_i) = \Pr(Z_i^* > 0) = \Pr(\beta X_i + \varepsilon_i > 0) = \Pr(-\varepsilon_i < \beta X_i) = F(\beta X_i) \quad (2)$$

where  $F$  is the cumulative distribution function of  $-\varepsilon_i$  [42]. Depending on the assumptions

regarding the functional form of  $F$ , probit or logit models can be used to model the determinants of digital service adoption.

### 3.2 Modelling impact

As explained, using digital extension services is expected to affect input use intensity, crop productivity, and income. Similar to Becerril and Abdulai [43] and Ogotu et al. [34], we link the adoption decision to the outcome variables by considering a simple model where a risk-neutral farmer maximizes income subject to a competitive output and input market and a single production function  $Q(W, X)$  that is continuous, strictly increasing, and strictly quasi-concave in a vector of variable inputs  $W$  and farm and household characteristics  $X$ . The household's income function can be represented as:

$$\max Y = PQ(W, X) - I W, \text{ subject to } Q(W, X) \geq Q \quad (3)$$

where  $Y$  is crop income,  $P$  is the output market price, and  $Q$  is the expected crop output quantity.  $I$  is a column vector of input prices, and  $W$  is a vector of input quantities. Further, the crop income function can also be expressed as a function of adopting digital extension services  $D$ , as well as market output and input prices and farm and household characteristics:

$$Y = f(D, I, P, X) \quad (4)$$

Thus, Eq (3) can be rewritten as:

$$\max Y(D, I, P, X) = PQ(W, X) - IW, \text{ subject to } Q(W, X) \geq Q \quad (5)$$

Now, applying Hotelling's lemma with respect to input and output prices, output supply and input demand can be obtained by simple differentiation, such that:

$$\frac{dY}{dI} = -W = W(D, I, P, X) \quad (6)$$

$$\frac{dY}{dP} = Q = Q(D, I, P, X) \quad (7)$$

From Eqs (6) and (7), it can be observed that a farm household's demand for inputs and the levels of crop output and income is influenced by the decision to adopt digital extension services, input, and output prices, as well as farm and household characteristics.

A common approach to estimate these relationships and the effect of digital extension services would be a set of regression models of the following type:

$$L_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \alpha_3 C + \mu_i \quad (8)$$

where  $L_i$  is the outcome variable of interest,  $C$  is a vector of relevant controls, including input and output prices, and  $\mu_i$  is a random error term. To evaluate how digital extension services are associated with the outcome variables, the coefficient  $\alpha_1$  is of particular interest. However, estimating Eq (8) will likely generate biased estimates of  $\alpha_1$  because farmers self-selected into adopting digital extension services, which may mean that  $D_i$  is correlated with the error term. As an alternative, we use propensity score matching (PSM) combined with several robustness checks. Further details are explained below.

### 3.3 Propensity score matching approach

A propensity score is the conditional probability of assignment to a particular treatment—adoption of digital extension services in our case—given a vector of observed covariates [44]. It

can be specified as:

$$p(X) = \Pr[D = 1|X] = E[D|X]$$

$$p(X) = F\{h(X_i)\} \quad (9)$$

where  $F\{\cdot\}$  is a normal or logistic cumulative distribution function and  $X$  is a vector of observed covariates [43].

The PSM method builds on two assumptions, namely the conditional independence assumption (CIA), which requires that outcome variables be independent of treatment conditional on the propensity score [45], and the presence of common support, which requires that treatment participants have comparable participants in the control group in terms of their propensity scores [46]. If these two conditions hold, the PSM estimator for the average treatment effect on the treated (ATT) can be specified as the mean difference in the outcome variable  $L$  over the common support, weighting the comparison units by the propensity score distribution of the participants as follows [47]:

$$ATT = E[L_{i1} - L_{i0}|D = 1]$$

$$ATT = E\{E[L_{i1} - L_{i0}|D_i = 1, p(X)]\}$$

$$ATT = E_{p(X)|D=1}\{E[L_{i1}|D_i = 1, p(X)] - E[L_{i0}|D_i = 0, p(X)]\}$$

where  $L_{i1}$  denotes the outcome of households adopting and  $L_{i0}$  the outcome of households not adopting digital extension services. We are using the term ATT, which is very common in the PSM literature, even though we are not claiming to identify strictly causal effects. While we try to reduce issues of endogeneity to the extent possible, establishing causality is difficult with cross-section observational data.

Successful matching first requires choosing a set of covariates that satisfy CIA, as omitting important variables can lead to biased estimates [48]. We use a logit regression with a large number of covariates chosen based on economic theory and past literature to estimate the propensity scores of treatment and control group participants.

Nevertheless, it is still possible that unobserved factors such as personal motivation, risk preferences, or entrepreneurial skills affect treatment assignments such that CIA would not hold. In order to reduce the risk of biased estimates due to relevant unobserved factors, we use an approach similar to Meemken and Qaim [49]. As part of the survey, we conducted a hypothetical bidding game to elicit respondent's willingness to pay (WTP) for an agri-tech platform that provides digital extension and improved access to input and output markets. After explaining the functioning of such an agri-tech platform, respondents were asked to quote the maximum price they would be willing to pay for a service that enables them to receive crop-related advisory information, order inputs that get delivered to their village, and find buyers for their output using a mobile application. The stated WTP is likely correlated with farmers' motivation, preferences, and entrepreneurial skills, so that it may be a good proxy of relevant unobserved characteristics [49–51]. Hence, including this WTP as an additional covariate when estimating the propensity scores can help to reduce potential issues of unobserved heterogeneity.

After estimating the propensity scores, we match treatment and control group farmers using three different matching algorithms, namely nearest neighbour matching (NNM), radius matching (RM), and kernel-based matching (KBM). Using and comparing different matching algorithms is common as a robustness check. In NNM, each treated individual is matched



with the three closest control group individuals in terms of their propensity scores. Here, matching is done with replacement, meaning that each control group individual can be used more than once as a match. For RM, a tolerance level on the maximum propensity score distance (caliper) is imposed. Here, the size of the caliper is defined as 0.2 of the standard deviation of the logit of the propensity score, as suggested by Austin [52]. KBM is a non-parametric matching estimator that uses weighted averages of all individuals in the control group to construct the counterfactual outcome [45,46]. After matching treatment and control group individuals based on their propensity scores, the ATT is calculated, as explained above.

Since we are estimating associations between digital extension services and several outcome variables, it is possible that some significant results occur simply due to chance meaning that false positives may be an issue. To address this potential issue when testing multiple hypotheses, we follow Arouna et al. [53] and present  $p$ -values corrected for the family-wise error rate (FWER), which is the probability of making at least one false discovery among a family of comparisons, using the Bonferroni correction and Holm [54] adjusted  $p$ -values (see S4 Table in the supporting information). For our main results, we also present sharpened  $q$ -values adjusted for multiple hypotheses testing [55].

Recent research indicated that using propensity scores for matching may lead to biased impact estimates in some situations [56]. To reduce the likelihood of bias, we use several alternative methods as robustness checks. First, we re-estimate all ATTs using inverse-probability weighted regression adjustment (IPWRA) [57]. Second, we estimate the regression models in Eq (8) using ordinary least squares (OLS) and including WTP as an additional regressor. As explained, our WTP variable is likely correlated with relevant unobserved characteristics, so that including WTP can reduce issues of unobserved heterogeneity. This approach has become popular in recent impact studies, especially in situations where valid instruments are hard to identify [50,58].

Third, to analyse the magnitude of possible omitted variable bias we calculate bias-adjusted coefficients using Oster bounds [59]. Fourth, we use coarsened exact matching (CEM) developed by Iacus et al. [60,61]. Unlike PSM, CEM bounds the maximum imbalance between treated and control groups *ex ante* based on the chosen level of coarsening. CEM also eliminates the need to restrict data to common empirical support and bounds both the error in estimating the average treatment effect and the amount of model dependence [60]. The idea of CEM is to temporarily coarsen each variable into meaningful groups and then match households *exactly* on each of their observed characteristics. Once one-to-one exact matching is done using the coarsened data, a simple difference in means between the outcome variables in the treated and control group can be used to estimate the causal effect, or, if the match is not exact, a linear regression model can be used to control for differences [62].

## 4. Results and discussion

### 4.1 Descriptive statistics

Our analysis is carried out with a total of 1028 farm household observations for which complete data for all relevant variables are available. Descriptive statistics of all outcome and explanatory variables for the full sample are presented in S3 Table in the supplementary information. The average age of the household head is 51 years. The average educational level is around 7 years of schooling. Most of the households are headed by a male; only 7% are headed by a female. Around 82% of the households belong to socially disadvantaged groups, including Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Classes (OBC). SC, ST, and OBC are among the most disadvantaged socioeconomic groups in India, recognized as needing special policy attention by the Indian Government. As is common in empirical

analysis with micro-level data from India, we use binary variables for these groups to account for differences in socioeconomic status.

In terms of farm sizes, the average household owns around 1.3 acres of land, even though the operational holding is larger with about 4.7 acres. In other words, many households lease in land from other landowners. Based on the operational land holding, 63% of the sample farmers are classified as marginal (<2.5 acres) or small (2.5–5 acres), and 37% are classified as medium (5–10 acres) or large (>10 acres). Around 50% of the cropped area is irrigated. On average, households grow 7 different crop species. In terms of commercialization levels, households sell around 43% of their crop output and travel about 5 km on average to the closest input and output markets. [S3 Table](#) also shows that 43% of the sample households had adopted digital extension services at the time of the survey, while 57% had not. [Table 1](#) shows the different types of information that digital extension adopters used.

[Table 2](#) presents descriptive statistics, disaggregated by digital extension service adopters and non-adopters. On average, adopting households cultivate somewhat more land, have older heads that are more likely to be male and are better educated than non-adopting households. While many of the differences in [Table 2](#) are statistically significant, adopting and non-adopting farms and households are still similar, so that comparative evaluation after controlling for observed differences seems justified. Looking at the outcome variables, which are shown in the lower part of [Table 2](#), adopters of digital extension services have more diversified cropping patterns, are more commercialized, and have higher crop incomes than non-adopters.

## 4.2 Factors influencing the adoption of digital extension services

Results of the logit model to explain the factors influencing digital extension service adoption are shown in [Table 3](#). Age, education, size of the land operated, asset ownership, and several other farm and household characteristics have a positive influence on the decision to adopt digital extension services. We use education and mobile phone ownership as proxies for digital literacy. Measuring digital literacy more directly with one variable or index is not straightforward, apart from the fact that such a variable in our context might possibly be associated with reverse causality issues. Age of the household head and size of the land operated have a non-linear influence on adoption. As the age of the household head increases, initially, there is a positive effect, but this positive effect becomes smaller with further increasing age. Similarly, as the size of the land operated increases, at first, there is a positive effect on digital extension service adoption, but this positive effect gets smaller with further increasing land size. The turning points for age and the size of the land operated are 45.7 years and 11.6 acres, respectively.

The WTP variable, which proxies for unobserved factors such as motivation, preferences, and entrepreneurial skills, is also positively associated with the adoption of digital extension services, as one would expect. Further, the size of the social network has a positive effect: a

**Table 1. Types of information used by digital extension service adopters.**

Type of information used	Percentage of adopters
Types of crops to grow	88%
Methods of cultivating selected crops	88%
Types of inputs to use	85%
Quantity of inputs to use	62%
Where to sell output	17%
Price to sell outputs	5%

<https://doi.org/10.1371/journal.pone.0259319.t001>

**Table 2. Socioeconomic characteristics of adopters and non-adopters of digital extension services.**

	Adopters		Non-adopters		Parametric tests <sup>a</sup>		Non-parametric tests <sup>b</sup>
	Mean	SD	Mean	SD	Difference	SE	p-value
Age of household head (years)	51.53	11.67	49.73	14.51	1.81**	(0.84)	0.021
Male household head (dummy)	0.96	0.20	0.92	0.27	0.04**	(0.02)	0.000
Household head owns a mobile phone (dummy)	0.76	0.43	0.70	0.46	0.06**	(0.03)	0.008
Illiterate: highest education of adult male (dummy)	0.05	0.23	0.10	0.29	-0.04**	(0.02)	0.043
Primary school: highest education of adult male (dummy)	0.21	0.41	0.27	0.45	-0.06**	(0.03)	0.017
Secondary school: highest education of adult male (dummy)	0.44	0.50	0.41	0.49	0.02	(0.03)	0.216
Bachelor or Masters: highest education of adult male (dummy)	0.29	0.46	0.19	0.39	0.11***	(0.03)	0.000
Scheduled tribe (dummy)	0.12	0.32	0.17	0.38	-0.05**	(0.02)	0.002
Scheduled caste (dummy)	0.12	0.32	0.21	0.41	-0.09***	(0.02)	0.000
Other backward classes (dummy)	0.56	0.50	0.44	0.50	0.12***	(0.03)	0.000
General caste (dummy)	0.21	0.41	0.18	0.38	0.03	(0.02)	0.309
Household size (number)	3.89	1.40	3.64	1.45	0.25***	(0.09)	0.000
Operated land (acres)	5.53	3.98	4.21	3.99	1.32***	(0.25)	0.000
Irrigation ratio (%)	53.98	35.75	48.49	38.56	5.49**	(2.36)	0.003
Livestock ownership (livestock units)	1.45	1.41	1.07	0.95	0.38***	(0.07)	0.000
Average distance to input and output market (km)	5.68	3.95	4.54	4.14	1.14***	(0.26)	0.000
Willingness to pay for digital agri-tech platform services (Rupees)	256.79	423.26	192.19	374.31	64.61***	(24.96)	0.000
Peer group <sup>c</sup>	13.91	8.47	11.61	9.75	2.30***	(0.58)	0.000
Off farm income (dummy)	0.60	0.49	0.69	0.46	-0.09***	(0.03)	0.002
<b>Outcome variables</b>							
Number of crops grown	8.45	4.83	6.35	4.41	2.10***	(0.29)	0.000
Seed expenditure (1,000 Rupees/acre) <sup>d</sup>	0.79	0.86	0.73	0.98	0.060	(0.06)	0.000
Fertilizer expenditure (1,000 Rupees/acre) <sup>d</sup>	1.80	1.35	1.73	1.62	0.07	(0.09)	0.005
Pesticides expenditure (1,000 Rupees/acre) <sup>d</sup>	0.69	0.73	0.65	0.95	0.03	(0.054)	0.000
Input expenditure (1,000 Rupees/acre) <sup>d</sup>	3.27	2.53	3.11	3.17	1.67	(1.83)	0.001
Crop productivity (1,000 Rs/acre) <sup>d</sup>	16.03	18.78	14.41	14.10	16.22	(1.02)	0.010
Commercialization (share of farm output sold 0–1)	0.51	0.28	0.38	0.33	0.13***	(0.02)	0.000
Crop income (1,000 Rs/acre) <sup>d</sup>	46.89	83.37	27.12	48.76	19.77***	(4.15)	0.000
Observations	440		588		1028		

## Notes

<sup>a</sup> *t*-test and *prtest* used for differences in means and proportions, respectively.

<sup>b</sup> Mann-Whitney test and Fisher exact test used for continuous and nominal variables, respectively.

<sup>c</sup> Number of households within the village from the same caste who adopted digital extension services.

<sup>d</sup> These monetary variables are used in logarithmic form in the regression models.

\* Significant at 10% level

\*\* Significant at 5% level

\*\*\* Significant at 1% level.

<https://doi.org/10.1371/journal.pone.0259319.t002>

larger number of people who adopted digital extension services from the same caste and living in the same village as the respondent is associated with a higher probability of individual adoption. In contrast, having off-farm income negatively influences the decision to adopt digital extension services (Table 3), probably because households pursuing off-farm income concentrate less on improving their farming business than households for whom agriculture is the only source of income. Off-farm income can have positive effects on farm investments, especially when access to agricultural credit is constrained [63]. However, the off-farm economy

**Table 3. Logit estimates of the propensity to adopt digital extension services.**

	<b>Coefficient</b>	<b>Robust SE</b>
Age of household head (years)	0.183***	(0.038)
Age squared	-0.002***	(0.000)
Male household head (dummy)	0.384	(0.329)
Household head owns a mobile phone (dummy)	0.332*	(0.191)
Primary school (dummy)	0.492*	(0.295)
Secondary school (dummy)	0.604**	(0.275)
Bachelor or Masters (dummy)	0.748**	(0.298)
Scheduled tribe (dummy)	-0.350	(0.288)
Scheduled caste (dummy)	-0.366	(0.302)
Other backward classes (dummy)	-0.189	(0.234)
Household size (number)	0.095*	(0.053)
Operated land (acres)	0.139***	(0.051)
Square of operated land (acres)	-0.006**	(0.003)
Irrigation ratio (%)	0.001	(0.002)
Livestock ownership (livestock units)	0.185**	(0.076)
Distance to input and output market (km)	0.028	(0.021)
WTP for digital agri-tech platform services (log)	0.295***	(0.093)
Peer group	0.031**	(0.013)
Off farm income (dummy)	-0.444***	(0.162)
Constant	-8.912***	(1.283)
Village dummies	Yes	
Observations	1028	
Log-likelihood	-580.957	
Pseudo R <sup>2</sup>	0.172	
<i>p</i> -value	0.0000	

\* Significant at 10% level

\*\* Significant at 5% level

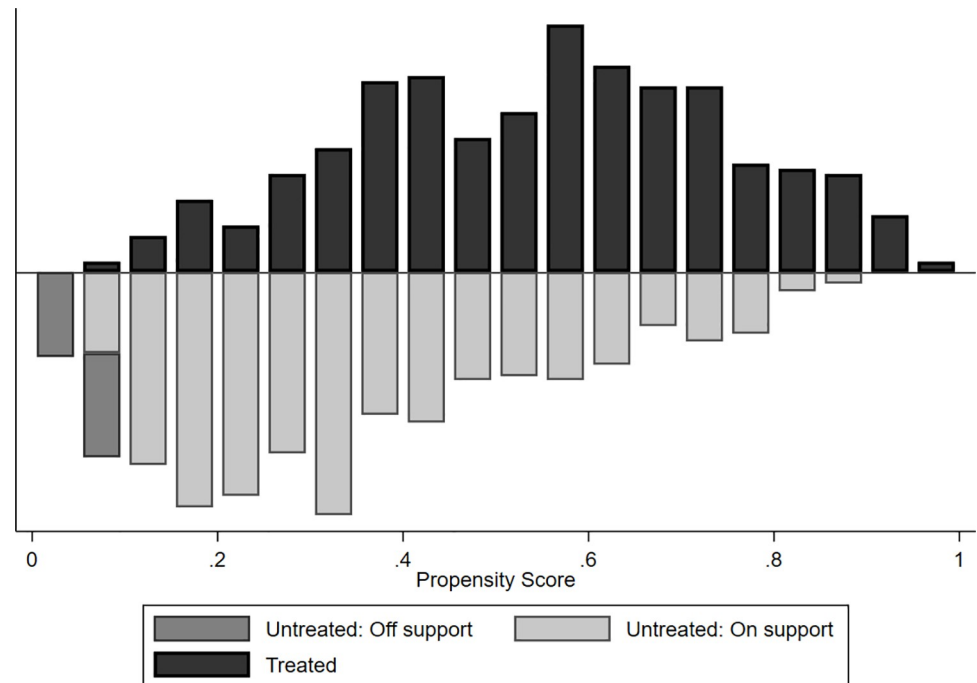
\*\*\* Significant at 1% level. Robust standard errors in parentheses.

<https://doi.org/10.1371/journal.pone.0259319.t003>

can be very diverse. For many farmers with very small landholdings, pursuing off-farm economic activities is often a simple survival mechanism rather than a conscious strategy to accumulate capital for farm upgrading [64].

In the logit model, we also control for distance to input and output markets. Market distance may be less relevant for the adoption of digital extension services that are offered in the local context but may certainly influence several of our outcome variables—such as input use and crop productivity—through various channels. Hence, it is important to control for general market access when calculating the propensity scores. As explained above, farmers do not operate the internet-based extension application themselves but through the FPO head. Hence, characteristics of the FPO head (motivation, communication skills, etc.) can also influence digital extension adoption and impacts. However, as our data were all collected in the same FPO with only one FPO head, controlling for FPO head characteristics is neither possible nor required to avoid bias.

Based on these logit estimates, we calculated the propensity scores for matching treatment group households that adopted digital extension services and control group households that did not adopt. Fig 1 presents the distribution of propensity scores and the region of overlap



**Fig 1. Distribution of estimated propensity scores and region of common support.**

<https://doi.org/10.1371/journal.pone.0259319.g001>

and common support. As can be seen, there is a substantial overlap of the propensity scores of both groups, meaning that the assumption of common support is satisfied.

### 4.3 Covariate balancing tests

Table 4 presents results of the covariate balancing tests to assess how well the propensity score matching performed, or, in other words, whether digital extension service adopters and non-adopters are comparable in terms of observed covariates. For the balancing tests, we calculated mean covariate differences and conducted independent sample *t*-tests before and after matching. A match is considered successful when all *t*-tests result in non-significant differences between the treatment and the control group after matching. The results in Table 4 confirm that successful matching is achieved with all three matching algorithms.

Another way of testing covariate balancing is through the standardized mean bias before and after matching. These results are presented in the lower part of Table 4. As can be seen, significant mean bias reduction is achieved with all three matching algorithms. A third diagnostic test is a comparison of the pseudo- $R^2$  from the logit model before and after matching, which is also shown in the lower part of Table 4. For all three matching algorithms, the pseudo- $R^2$  after matching is low, suggesting that the systematic differences in the covariates that existed before matching were successfully removed.

### 4.4 Association between digital extension services and agricultural performance

Table 5 presents the PSM treatment effects, showing how the adoption of digital extension services is associated with agricultural performance. All three matching estimators indicate that adopting digital extension services is positively associated with all indicators of agricultural performance. After controlling for confounding factors, digital extension adopters cultivate

Table 4. Covariate balancing tests.

Covariates	Nearest neighbour matching			Radius matching			Kernel matching		
	% Bias reduction	<i>p</i> -value mean difference, unmatched	<i>p</i> -value mean difference, matched	% Bias reduction	<i>p</i> -value mean difference, unmatched	<i>p</i> -value mean difference, matched	% Bias reduction	<i>p</i> -value mean difference, unmatched	<i>p</i> -value mean difference in, matched
Age of household head	82.2	0.035	0.727	78.0	0.035	0.633	77.1	0.035	0.621
Age squared	82.8	0.226	0.848	54.7	0.226	0.576	54.9	0.226	0.578
Male household head	82.1	0.000	0.434	96.2	0.000	0.875	99.5	0.000	0.983
Owns mobile phone	69.0	0.007	0.481	86.5	0.007	0.731	90.7	0.007	0.813
Primary school	28.2	0.016	0.160	93.9	0.016	0.891	90.1	0.016	0.825
Secondary school	89.3	0.198	0.915	88.1	0.198	0.893	84.5	0.198	0.861
Bachelor or Master	66.9	0.000	0.355	88.8	0.000	0.721	82.8	0.000	0.580
Scheduled tribe	100.0	0.002	1.000	90.6	0.002	0.750	92.6	0.002	0.801
Scheduled caste	70.0	0.000	0.290	88.7	0.000	0.644	86.7	0.000	0.588
Other backward classes	89.4	0.000	0.710	99.8	0.000	0.994	97.2	0.000	0.912
Household size	66.7	0.001	0.421	93.2	0.001	0.849	89.4	0.001	0.765
Operated land	71.3	0.000	0.188	97.8	0.000	0.910	94.5	0.000	0.775
Square of operated land	35.3	0.005	0.135	90.1	0.005	0.791	86.4	0.005	0.717
Irrigation ratio	63.0	0.021	0.483	80.9	0.021	0.686	76.0	0.021	0.612
Livestock ownership	72.5	0.000	0.180	90.5	0.000	0.588	94.8	0.000	0.767
Distance input/ output market	69.3	0.000	0.281	96.8	0.000	0.899	89.9	0.000	0.686
WTP (log)	67.2	0.000	0.128	94.6	0.000	0.767	89.6	0.000	0.569
Peer group	85.3	0.000	0.603	84.6	0.000	0.528	88.8	0.000	0.646
Off-farm income	44.3	0.002	0.202	96.0	0.002	0.915	98.5	0.002	0.967
Mean bias before matching		16.7			16.7			16.7	
Mean bias after matching		5.3			2.2			2.5	
<i>p</i> -value of LRChi <sup>2</sup> unmatched		0.000			0.000			0.000	
<i>p</i> -value of LRChi <sup>2</sup> matched		0.954			1.000			1.000	
Pseudo-R <sup>2</sup> unmatched		0.174			0.174			0.174	
Pseudo-R <sup>2</sup> matched		0.027			0.005			0.006	

<https://doi.org/10.1371/journal.pone.0259319.t004>

one more crop species on their farm than non-adopters. The monetary outcome variables are log-transformed, so the respective ATTs can be interpreted in percentage terms. The results in Table 5 suggest that digital extension adopters have 15–20% higher input intensity, around 18% higher crop productivity, and 25–29% higher crop incomes than non-adopters. Adopters

**Table 5. Relationship between adoption of digital extension services and agricultural performance (PSM results).**

Outcome variable	Nearest neighbour matching		Radius matching		Kernel matching	
	ATT	SE	ATT	SE	ATT	SE
Number of crops grown	1.211*** [0.024]	(0.443)	1.017*** [0.019]	(0.371)	1.095*** [0.006]	(0.355)
Seed expenditure per acre (log)	0.170 [0.153]	(0.115)	0.200** [0.046]	(0.099)	0.198** [0.035]	(0.097)
Fertilizer expenditure per acre (log)	0.161** [0.039]	(0.077)	0.149** [0.032]	(0.062)	0.153** [0.028]	(0.064)
Pesticide expenditure per acre (log)	0.199** [0.047]	(0.102)	0.195** [0.032]	(0.086)	0.198** [0.028]	(0.083)
Total expenditure per acre (log)	0.188** [0.028]	(0.079)	0.194*** [0.012]	(0.066)	0.197*** [0.006]	(0.066)
Crop productivity (log)	0.175** [0.024]	(0.065)	0.175*** [0.032]	(0.059)	0.177*** [0.006]	(0.058)
Crop commercialization	0.074*** [0.024]	(0.028)	0.049** [0.046]	(0.024)	0.049** [0.035]	(0.023)
Crop income (log)	0.285** [0.039]	(0.132)	0.254** [0.046]	(0.107)	0.265** [0.024]	(0.107)

ATT: average treatment effect on the treated. PSM: propensity score matching. Bootstrapped standard errors with 1,000 replications are shown in parentheses.

\* Significant at 10% level

\*\* Significant at 5% level

\*\*\* Significant at 1% level. Multiple hypotheses corrected sharpened  $q$ -values following Anderson [54] are presented in square brackets. Unadjusted  $p$ -values and Bonferroni and Holm adjusted  $p$ -values are shown in S4 Table in the supplementary information. Results with bootstrapped standard errors and 10,000 replications are shown in S5 Table.

<https://doi.org/10.1371/journal.pone.0259319.t005>

are also significantly more commercialized (by 5–7 percentage points) than non-adopters. These results remain statistically significant also after correcting for multiple hypothesis testing (Table 5).

Several of the variables used to calculate the propensity scores may potentially be endogenous. This is particularly true for mobile phone ownership, off-farm income, and the peer group variable. Using endogenous variables for propensity score calculations is unproblematic if these are not affected by the “treatment” [45,65]. However, as reverse causality cannot be ruled out with certainty, we tested how the ATT estimates change when we exclude these potentially endogenous variables in the logit regression used to calculate the propensity scores. The ATTs are very similar to the original ones (see S6 Table in the supporting information),

**Table 6. Critical level of hidden bias ( $\Gamma$ ).**

	Nearest neighbour matching	Radius matching	Kernel matching
Number of crops grown	1.30	1.15	1.20
Seed expenditure per acre (log)	1.15	1.45	1.40
Fertilizer expenditure per acre (log)	1.30	1.40	1.45
Pesticide expenditure per acre (log)	1.30	1.45	1.45
Total input expenditure per acre (log)	1.35	1.60	1.65
Crop productivity (log)	1.40	1.50	1.50
Crop commercialization	1.40	1.30	1.30
Crop income (log)	1.30	1.45	1.50

<https://doi.org/10.1371/journal.pone.0259319.t006>

thus underlining the robustness of the estimates. Another variable that may potentially suffer from reverse causality is WTP, our proxy of potentially relevant unobserved heterogeneity. We also tried excluding WTP along with the other potentially endogenous variables, with results shown in [S7 Table](#) in the supplementary information. These results are slightly different from the previous estimates, as one would expect, but most of the ATTs remain statistically significant, which is reassuring.

Another concern may be related to the fact that we express most of the outcome variables in logarithmic form, even though several include zero or—in the case of crop income—even negative observations (see [S3 Table](#)), for which the logarithm is not defined. It should be noted that the number of households with zero or negative observations for the outcome variables is small and mostly confined to the group of non-adopters. Very few of these households with zero or negative observations were selected as relevant matches based on their propensity scores. Nevertheless, we reran the PSM estimates also in linear form, without taking logs, and obtained similar results. For some of the outcome variables, the estimated ATTs even increased (see [S8 Table](#) in the supplementary information). With the outcome variables expressed in linear form, the digital extension effect on crop productivity increased to 21%, and the effect on crop income increased to around 40%. Hence, using the log-transformation seems to lead to conservative estimates.

We cautiously conclude that digital technologies that use data from farms to provide personalized information are effective in terms of helping farmers to make better cropping, technology, and input decisions. The information provided by the digital extension services on the best types of crops to grow, the appropriate methods of cultivation, the optimal timing of the different operations, and the suitable input regimes seems to help farmers allocate their resources more efficiently. This leads to the use of better technologies and higher input intensities, thus increasing crop productivity and income.

#### 4.5 Sensitivity to potential hidden bias

The PSM estimates of the ATTs may still be biased if any unobserved variables affect treatment assignment and are also correlated with the outcome variables. Using the PSM estimates, we analyse how sensitive our results are to such potential hidden bias by calculating Rosenbaum bounds [66]. Rosenbaum bounds estimate critical levels of hidden bias ( $\Gamma$ ) for the ATTs at which the conclusion of a positive treatment effect would have to be challenged. The results are shown in [Table 6](#).

The robustness of the ATTs to potential hidden bias varies across the outcome variables and the three matching methods ([Table 6](#)). A value of 1.30 for the number of crops grown and the nearest neighbour matching algorithm means that the respective ATT would remain positive and significant at the 90% level even if there were hidden bias up to a magnitude of 30% (meaning 30% systematic difference between treatment and control group in terms of unobserved factors even after matching). Only if there were hidden biases of more than 30%, the ATT would turn insignificant. Remember that when estimating propensity scores for matching we did not only include many farm and household covariates but also WTP as a proxy for relevant unobserved factors. Against this background, we expect that any remaining hidden bias would be lower than 30%. For many of the other outcome variables, the critical levels for hidden bias are larger than 1.30 ([Table 6](#)), meaning that the conclusion of significantly positive associations is robust.

#### 4.6 Additional robustness checks

To further test the robustness of the PSM estimates, we used several alternative methods, as explained in section 3.3. The ATTs obtained with the IPWRA method are shown in [S9 Table](#)



in the supplementary information. They are very similar to those obtained with PSM in terms of both their size and significance levels. The OLS results with the inclusion of WTP to control for unobserved heterogeneity are shown in [S10 Table](#). These results are also very similar to both the PSM and IPWRA results.

Further, to show that our results are not driven by omitted variable bias we estimate Oster bounds using multiple specifications. To calculate the bounds to the true value of the treatment effect, we need to fix two parameters, namely (i)  $\delta$ , which indicates the relative importance of unobservables compared to observables, and (ii)  $R_{\max}$ , which is the maximum R-squared that can be achieved from a hypothetical regression that controls for all variables including unobserved factors. Oster [59] argues that  $\tilde{R} < R_{\max} < 1$ , where  $\tilde{R}$  is the R-squared from the estimated equation. The standard case in the literature is to use  $R_{\max}$  equal to 1.3 times the R-square value of the estimated equation with full controls [59,67–69]. We follow this convention and present results with varying assumptions for  $\delta$  in [S10 Table](#). In all cases, the treatment bounds do not include the value zero, except for crop income when large values for  $\delta$  are assumed. Overall, these results are reassuring.

Finally, we present results of coarsened exact matching (CEM) in [S12](#) and [S13](#) Tables in the supplementary information. Here, we use a smaller set of covariates for matching, as the inclusion of many variables leads to fewer exact matches. Instead, we control for remaining imbalances after matching using linear regression models. For CEM, we coarsen continuous variables such as age, household size, and land size, into sub-categories. Age is coarsened into bins of (20–40), (40–60), and (above 60 years) to categorize household heads as young adult, middle aged, and older heads respectively. The binning strategy for operated land is small farms (2.5–5 acres), medium farms (5–10 acres), and large farms (>10 acres), and the binning strategy for household size is 1(0–4 household members) and 0 (more than 4 members). The remaining variables were already in dichotomous form, so we use those variables as in the original data. Each household is assigned a binning signature based on the coarsened and dichotomous variables and then exactly matched with households in the other group. The ATT estimation results are shown in [S13 Table](#). They are very similar to the PSM results discussed above.

Based on the various robustness checks, we conclude that our findings are quite robust to variations in matching approaches, estimation methods, and possible unobserved heterogeneity.

#### 4.7 Limitations

While we have tried hard to control and test for potential biases, several limitations need to be pointed out. First, our analysis relies on cross-section observational data where the establishment of causality is difficult. Hence, our results should be interpreted as associations between digital extension adoption and agricultural performance and not as fully identified net effects. Although we addressed issues of endogeneity to the extent possible, follow-up research with panel data and/or experimental approaches could be useful to further improve the identification strategy. Second, the results from one example of an agri-tech platform in one farmer organization in eastern India should not be generalized. Additional studies in other contexts would be useful to increase the external validity of the results. Third, we concentrated on a few outcome variables related to crop production and crop income, as this is what the agri-tech platform in the study region focuses on. Crop productivity and crop income are not comprehensive measures of household welfare. Future studies could analyse other important outcomes related to total household income, food security, time allocation, and gender roles, among others. Fourth, we looked at effectiveness in terms of improving agricultural

performance without considering the costs of providing and using digital services. Studies on cost-effectiveness would be useful to gain further policy-relevant insights.

## 5. Conclusion

Traditionally, the diffusion of agricultural information in developing countries has been promoted through the public extension service, wherein extension agents visit and educate individual farmers or farmer groups. This traditional way of information dissemination has two major drawbacks. First, as personal visits are associated with high transaction costs, only a very limited number of farmers can be reached. Second, the information provided through this channel is often generic and not necessarily well adapted to farmers' specific needs and conditions. Using digital approaches and technologies can potentially improve the effectiveness of agricultural extension services by reducing transaction costs and improving the quality of the information provided. Data-driven algorithms can be used to tailor information to the specific conditions of individual farmers. However, research on the actual effectiveness of such personalized digital extension approaches in the small farm sector is still very limited.

In this study, we collected and used data from smallholder vegetable farmers in eastern India to analyse the association between using personalized digital extension services and agricultural performance. The digital agri-tech platform that had recently been launched in the study region provides advice on which types of crops to grow and inputs to use, considering the weather, soil, and other agronomic as well as socioeconomic conditions of each farmer. For the analysis, we looked at various outcome variables such as crop diversity, input intensity, crop productivity, levels of commercialization, and crop income.

As farmers decided themselves whether to adopt the digital extension services, the evaluation needed to deal with possible issues of selection bias. We employed propensity score matching to reduce bias due to observed heterogeneity between digital extension service adopters and non-adopters. Moreover, we used a willingness to pay (WTP) variable as an additional covariate in the propensity score model to also reduce possible bias from unobserved heterogeneity. As robustness checks, we used and compared different matching methods (PSM and CEM), employed other econometric tools (OLS and IPWRA), and corrected for multiple hypothesis testing.

Our results show that adopting personalized digital extension services is positively associated with several of the agricultural performance indicators. After controlling for confounding factors, digital extension adopters were found to have 15–20% higher input intensity, 18% higher crop productivity, and 18–29% higher crop incomes than non-adopters. These findings suggest that digital approaches and technologies can be effective tools to improve personalized agricultural extension and promote smallholder productivity and income.

The associations identified here are relatively large in magnitude and larger than what several other studies on digital agricultural extension found [20,21,34]. Given that we cannot completely rule out any remaining bias, our associations may potentially be overestimated. On the other hand, as discussed above, most existing studies evaluated digital extension services providing generic advice, not personalized advice as analysed here. Personalized advice that is tailored to farmers' specific needs and conditions should be expected to have larger effects. The only existing studies evaluating the effects of personalized digital extension are Arouna et al. [38] and Oyinbo et al. [39]. Arouna et al. [38] found that personalized fertilizer recommendations provided through digital tools lead to yield increases of 7% and profit increases of 10% in Nigeria. Our estimates are larger, but the Farmex platform analyzed here is more comprehensive and provides digital advice not only on fertilizer use but on several other aspects as well, including crop choices and pest control, among others. Hence, larger estimates are not

unexpected, even though we stress that the exact magnitude of the associations found here should be interpreted with some caution. In any case, our results suggest that personalized digital extension services have promising potential to promote agricultural development in the small farm sector.

The agri-tech platform in the study region in India is still relatively simple and can be further improved. So far, the services provided concentrate only on the input side, although there are plans to extend the services and use the platform to also link farmers directly to retailers and consumers. Technological developments in areas such as open-source software and big data analytics may further improve the types of services provided. In the Indian example analysed here, the services are provided free of charge to farmers through a social business enterprise. However, if measurable benefits are confirmed, farmers may also be willing to pay a certain amount for such digital services, as indicated by significantly positive WTP estimates in our analysis.

Nevertheless, some public support may be needed to make digital extension services effective and accessible for a large number of farmers. For instance, certain infrastructure elements—such as roads, electricity, telephone network, and internet coverage—are important preconditions for digital service providers to become active in a region. In addition, a minimum level of computer and digital literacy is required either among farmers or at least among local intermediaries. Therefore, from a policy perspective, investments in rural road and ICT infrastructure, in promoting digital literacy among rural households, and in creating an enabling business environment for related entrepreneurial activities are important steps towards fostering agricultural innovation and equitable growth in the small-farm sector.

## Supporting information

### **S1 Data.**

(XLS)

### **S1 Dataset. Dataset dictionary.**

(XLS)

### **S1 Table. Number of farmers by village and sample size.**

(DOCX)

### **S2 Table. Summary statistics of control group farmers by FPO membership.**

(DOCX)

### **S3 Table. Summary statistics of full sample.**

(DOCX)

### **S4 Table. Correction for multiple inferences of the treatment effects.**

(DOCX)

### **S5 Table. PSM estimates with bootstrapped standard errors with 10,000 replications (robustness check).**

(DOCX)

### **S6 Table. PSM estimates excluding potentially endogenous variables (robustness check).**

(DOCX)

### **S7 Table. PSM estimates excluding WTP and other potentially endogenous variables (robustness check).**

(DOCX)

**S8 Table. PSM estimates with outcome variables in linear form.**

(DOCX)

**S9 Table. Association between adopting digital extension services and agricultural performance (IPWRA results).**

(DOCX)

**S10 Table. OLS estimates with WTP as additional control variable (robustness check).**

(DOCX)

**S11 Table. Oster bounds.**

(DOCX)

**S12 Table. LI measure of imbalance before and after coarsened exact matching.**

(DOCX)

**S13 Table. Estimated associations based on coarsened exact matched data.**

(DOCX)

## Acknowledgments

The authors thank Joachim von Braun and Zaneta Kubik for their valuable comments. They also thank Suvankar Mishra and local partners in India for their support during the fieldwork, as well as the farm households that kindly participated in the study. We also thank three anonymous reviewers of this journal for their constructive comments.

## Author Contributions

**Conceptualization:** Pallavi Rajkhowa.

**Data curation:** Pallavi Rajkhowa.

**Formal analysis:** Pallavi Rajkhowa.

**Investigation:** Pallavi Rajkhowa.

**Methodology:** Pallavi Rajkhowa, Matin Qaim.

**Software:** Pallavi Rajkhowa.

**Supervision:** Matin Qaim.

**Writing – original draft:** Pallavi Rajkhowa.

**Writing – review & editing:** Matin Qaim.

## References

1. Mellor JW, Malik SJ. The impact of growth in small commercial farm productivity on rural poverty reduction. *World Dev.* 2017 Mar 1; 91:1–10. Available from: <https://doi.org/10.1016/j.worlddev.2016.09.004>.
2. Ogutu SO, Qaim M. Commercialization of the small farm sector and multidimensional poverty. *World Dev.* 2019 Feb 1; 114:281–93. Available from: <https://doi.org/10.1016/j.worlddev.2018.10.012>.
3. Key N, Sadoulet E, de Janvry A. Transactions costs and agricultural household supply response. *Am J Agric Econ.* 2000; 82(2):245–59. Available from: <https://doi.org/10.1111/0002-9092.00022>.
4. Takahashi K, Muraoka R, Otsuka K. Technology adoption, impact, and extension in developing countries' agriculture: A review of the recent literature. *Agric Econ.* 2020 Jan 4; 51(1):31–45. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/agec.12539>.
5. Taylor M, Bhasme S. Model farmers, extension networks and the politics of agricultural knowledge transfer. *J Rural Stud.* 2018 Nov 1; 64:1–10. Available from: <https://doi.org/10.1016/j.jrurstud.2018.09.015>.

6. Aker JC. Dial "A" for agriculture: A review of information and communication technologies for agricultural extension in developing countries. *Agricultural Economics*. 2011; 42: 631–647. Available from: <https://doi.org/10.1111/j.1574-0862.2011.00545.x>.
7. Aker JC, Mbiti IM. Mobile phones and economic development in Africa. *J Econ Perspect*. 2010 Jun; 24(3):207–32. Available from: <https://www.aeaweb.org/articles?id=10.1257/jep.24.3.207>.
8. Jensen R. The digital divide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *Q J Econ*. 2007 Aug; 122(3):879–924. Available from: <https://doi.org/10.1162/qjec.122.3.879>.
9. Nakasone E, Torero M, Minten B. The power of information: The ICT revolution in agricultural development. *Annu Rev Resour Econ*. 2014; 6(1):533–50. Available from: <http://www.annualreviews.org/doi/10.1146/annurev-resource-100913-012714>.
10. Niebel T. ICT and economic growth—Comparing developing, emerging and developed countries. *World Dev*. 2018 Apr 1; 104:197–211. Available from: <https://doi.org/10.1016/j.worlddev.2017.11.024>.
11. Torero M, von Braun J. *Information and Communication Technologies for Development and Poverty Reduction: The Potential of Telecommunications*. Johns Hopkins University Press; 2006. 362 p.
12. Aker JC, Fafchamps M. Mobile phone coverage and producer markets: Evidence from West Africa. *World Bank Econ Rev*. 2014; 29(2):262–92. Available from: <https://doi.org/10.1093/wber/ltu006>.
13. Haile MG, Wossen T, Kalkuhl M. Access to information, price expectations and welfare: The role of mobile phone adoption in Ethiopia. *Technol Forecast Soc Change*. 2019 Aug 1; 145:82–92. Available from: <https://doi.org/10.1016/j.techfore.2019.04.017>.
14. Shimamoto D, Yamada H, Gummert M. Mobile phones and market information: Evidence from rural Cambodia. *Food Policy*. 2015 Nov 1; 57:135–41. Available from: <https://doi.org/10.1016/j.foodpol.2015.10.005>.
15. Fan Q, Salas Garcia VB. Information access and smallholder farmers' market participation in Peru. *J Agric Econ*. 2018 Jun 1; 69(2):476–94. Available from: <http://doi.wiley.com/10.1111/1477-9552.12243>.
16. Muto M, Yamano T. The impact of mobile phone coverage expansion on market participation: Panel data evidence from Uganda. *World Dev*. 2009 Dec 1; 37(12):1887–96. Available from: <https://doi.org/10.1016/j.worlddev.2009.05.004>.
17. Tadesse G, Bahiigwa G. Mobile phones and farmers' marketing decisions in Ethiopia. *World Dev*. 2015 Apr 1; 68:296–307. Available from: <https://doi.org/10.1016/j.worlddev.2014.12.010>.
18. Zanello G. Mobile phones and radios: Effects on transactions costs and market participation for households in Northern Ghana. *J Agric Econ*. 2012 Sep 1; 63(3):694–714. Available from: <http://doi.wiley.com/10.1111/j.1477-9552.2012.00352.x>.
19. Aker JC, Ksoll C. Can mobile phones improve agricultural outcomes? Evidence from a randomized experiment in Niger. *Food Policy*. 2016; 60:44–51. Available from: <https://doi.org/10.1016/j.foodpol.2015.03.006>.
20. Baumüller H. The little we know: An exploratory literature review on the utility of mobile phone-enabled services for smallholder farmers. *J Int Dev*. 2018 Jan 1; 30(1):134–54. Available from: <http://doi.wiley.com/10.1002/jid.3314>.
21. Fu X, Akter S. The impact of mobile phone technology on agricultural extension services delivery: Evidence from India. *J Dev Stud*. 2016 Nov 1; 52(11):1561–76. Available from: <https://doi.org/10.1080/00220388.2016.1146700>.
22. Parlasca MC, Mußhoff O, Qaim M. Can mobile phones improve nutrition among pastoral communities? Panel data evidence from Northern Kenya. *Agric Econ*. 2020 Mar; 51(3):475–88. Available from: <https://doi.org/10.1111/agec.12566>.
23. Sekabira H, Qaim M. Can mobile phones improve gender equality and nutrition? Panel data evidence from farm households in Uganda. *Food Policy*. 2017 Dec; 73:95–103. Available from: <https://doi.org/10.1016/j.foodpol.2017.10.004>.
24. Leng C, Ma W, Tang J, Zhu Z. ICT adoption and income diversification among rural households in China. *Appl Econ*. 2020; 52:33, 3614–3628. Available from: <https://doi.org/10.1080/00036846.2020.1715338>.
25. Muto M. The impacts of mobile phones and personal networks on rural-to-urban migration: Evidence from Uganda. *J Afr Econ*. 2012; 21(5):787–807. Available from: <https://doi.org/10.1093/jae/ejs009>.
26. Kikulwe EM, Fischer E, Qaim M. Mobile Money, Smallholder Farmers, and Household Welfare in Kenya. Mahal A, editor. *PLoS One*. 2014 Oct 6; 9(10):e109804. Available from: <https://dx.plos.org/10.1371/journal.pone.0109804>. <https://doi.org/10.1371/journal.pone.0109804> PMID: 25286032
27. Sekabira H, Qaim M. Mobile money, agricultural marketing, and off-farm income in Uganda. *Agric Econ*. 2017 Sep 1; 48(5):597–611. Available from: <http://doi.wiley.com/10.1111/agec.12360>.

28. Hübler M, Hartje R. Are smartphones smart for economic development? *Econ Lett*. 2016 Apr 1; 141:130–3. Available from: <https://doi.org/10.1016/j.econlet.2016.02.001>.
29. Ma W, Nie P, Zhang P, Renwick A. Impact of Internet use on economic well-being of rural households: Evidence from China. *Rev Dev Econ*. 2020 Feb 5;rode.12645. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/rode.12645>.
30. Nie P, Ma W, Sousa-Poza A. The relationship between smartphone use and subjective well-being in rural China. *Electron Commer Res*. 2020 Jan 14;1–27. Available from: <https://doi.org/10.1007/s10660-020-09397-1>.
31. Rao NC, Sutradhar R, Reardon T. Disruptive innovations in food value chains and small farmers in India. *Ind Jn Agri Econ*. 2017; 72(1). Available from: [http://www.iegindia.org/upload/profile\\_publication/171117\\_115238Rao et al., 2017.pdf](http://www.iegindia.org/upload/profile_publication/171117_115238Rao et al., 2017.pdf).
32. Fafchamps M, Minten B. Impact of SMS-based agricultural information on indian farmers. *World Bank Econ Rev*. 2012 Feb 27; 26(3):383–414. Available from: <https://doi.org/10.1093/wber/lhr056>.
33. Goyal A. Information, direct access to farmers, and rural market performance in central India. *Am Econ J Appl Econ*. 2010 Jul; 2(3):22–45. Available from: <https://www.aeaweb.org/articles?id=10.1257/app.2.3.22>.
34. Ogutu SO, Okello JJ, Otieno DJ. Impact of information and communication technology-based market information services on smallholder farm input use and productivity: The case of Kenya. *World Dev*. 2014; 64(104482):311–21. Available from: <http://dx.doi.org/10.1016/j.worlddev.2014.06.011>.
35. Omulo G, Kumeh EM. Farmer-to-farmer digital network as a strategy to strengthen agricultural performance in Kenya: A research note on 'Wefarm' platform. *Technol Forecast Soc Change*. 2020 Sep 1; 158:120120. Available from: <https://doi.org/10.1016/j.techfore.2020.120120>.
36. Van Campenhout B, Vandeveldel S, Walukano W, Van Asten P. Agricultural Extension Messages Using Video on Portable Devices Increased Knowledge about Seed Selection, Storage and Handling among Smallholder Potato Farmers in Southwestern Uganda. *PLoS One*. 2017 Jan 25; 12(1):e0169557. Available from: <https://doi.org/10.1371/journal.pone.0169557> PMID: 28122005
37. Van Campenhout B, Spielman DJ, Lecoutere E. Information and Communication Technologies to Provide Agricultural Advice to Smallholder Farmers: Experimental Evidence from Uganda. *Am J Agric Econ*. 2021 Jan 26; 103(1):317–37. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/ajae.12089>.
38. Arouna A, Michler JD, Yergo WG, Saito K. One Size Fits All? Experimental Evidence on the Digital Delivery of Personalized Extension Advice in Nigeria. *Am J Agric Econ*. 2020 Oct 16;ajae.12151. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/ajae.12151>. PMID: 33230345
39. Oyinbo O, Chamberlin J, Abdoulaye T, Maertens M. Digital Extension, Price Risk, and Farm Performance: Experimental Evidence from Nigeria. *Am J Agric Econ*. 2021; Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/ajae.12242>.
40. Fischer E, Qaim M. Linking Smallholders to Markets: Determinants and Impacts of Farmer Collective Action in Kenya. *World Dev*. 2012 Jun 1; 40(6):1255–68. Available from: <https://doi.org/10.1016/j.worlddev.2011.11.018>.
41. Carletto C, Jolliffe D, Banerjee R. From tragedy to renaissance: Improving agricultural data for better policies. *J Dev Stud*. 2015 Feb 1; 51(2):133–48. Available from: <https://doi.org/10.1080/002203882014968140>.
42. Cameron AC, Trivedi PK. *Microeconometrics: Methods and Applications*. Cambridge University Press; 2005. 1034 p.
43. Becerril J, Abdulai A. The impact of improved maize varieties on poverty in Mexico: A propensity score-matching approach. *World Dev*. 2010 Jul; 38(7):1024–35. Available from: <https://doi.org/10.1016/j.worlddev.2009.11.017>.
44. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *Am Stat*. 1985 Feb; 39(1):33. Available from: <https://www.jstor.org/stable/2683903?origin=crossref>.
45. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv*. 2008 Feb 1; 22(1):31–72. Available from: <http://doi.wiley.com/10.1111/j.1467-6419.2007.00527.x>.
46. Guo S, Fraser MW. *Propensity Score Analysis: Statistical Methods and Applications*. 2015. 421 p.
47. Khandker SR, Koolwal GB, Samad HA. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Vol. 41, The World Bank. 2010. 1–262 p.
48. Dehejia RH, Wahba S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *J Am Stat Assoc*. 1999 Dec 1; 94(448):1053–62. Available from: <https://www.tandfonline.com/action/journalInformation?journalCode=uasa20>.

49. Meemken E-M, Qaim M. Can private food standards promote gender equality in the small farm sector? *J Rural Stud.* 2018 Feb 1; 58:39–51. Available from: <https://doi.org/10.1016/j.jrurstud.2017.12.030>.
50. Bellemare MF, Novak L. Contract farming and food security. *Am J Agric Econ.* 2017 Mar 1; 99(2):357–78. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1093/ajae/aaw053>.
51. Verhofstadt E, Maertens M. Smallholder cooperatives and agricultural performance in Rwanda: do organizational differences matter? *Agric Econ.* 2014 Nov 1; 45(S1):39–52. Available from: <http://doi.wiley.com/10.1111/agec.12128>.
52. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat.* 2011; 10(2):150–61. Available from: <https://doi.org/10.1002/pst.433> PMID: 20925139
53. Arouna A, Michler JD, Lokossou JC. Contract farming and rural transformation: Evidence from a field experiment in Benin. *J Dev Econ.* 2021 Jun 1; 151:102626. Available from: <https://doi.org/10.1016/j.jdeveco.2021.102626> PMID: 34177043
54. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Stat.* 1979 Sep 12; 6(2):65–70. Available from: <http://www.jstor.org/stable/4615733>.
55. Anderson ML. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *J Am Stat Assoc.* 2008 Dec 1; 103(484):1481–95. Available from: <https://doi.org/https://doi.org/10.1198/016214508000000841>
56. King G, Nielsen R. Why Propensity Scores Should Not Be Used for Matching. *Polit Anal.* 2019 Oct 1; 27(4):435–54. Available from: <https://doi.org/10.1017/pan.2019.11>.
57. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data.* Vol. 1, MIT Press Books. 2010. 1064 p.
58. Ruml A, Qaim M. New evidence regarding the effects of contract farming on agricultural labor use. *Agric Econ.* 2021 Jan 12; 52(1):51–66. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/agec.12606>.
59. Oster E. Unobservable Selection and Coefficient Stability: Theory and Evidence. *J Bus Econ Stat.* 2019 Apr 3; 37(2):187–204. Available from: <https://doi.org/10.1080/07350015.2016.1227711>.
60. Iacus SM, King G, Porro G. Causal inference without balance checking: Coarsened exact matching. *Polit Anal.* 2017/01/04. 2012; 20(1):1–24. Available from: <https://doi.org/10.1093/pan/mpr013>.
61. Iacus SM, King G, Porro G. Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *J Am Stat Assoc.* 2011 Mar 1; 106(493):345–61. Available from: <https://doi.org/10.1198/jasa.2011.tm09599>.
62. Blackwell M, Iacus S, King G, Porro G. cem: Coarsened exact matching in Stata. *Stata J.* 2009; 9(4):524–46.
63. Haggblade S, Hazell PBR, Reardon T. *Transforming the rural nonfarm economy. Opportunities and threats in the developing world.* Johns Hopkins University Press, Baltimore; 2007.
64. Davis B, Winters P, Reardon T, Stamoulis K. Rural nonfarm employment and farming: household-level linkages. *Agric Econ.* 2009 Mar 1; 40(2):119–23. Available from: <https://doi.org/10.1111/j.1574-0862.2009.00374.x>.
65. Heckman JJ, Vytlacil E. Structural Equations, Treatment Effects, and Econometric Policy Evaluation1. *Econometrica.* 2005 May 1; 73(3):669–738. Available from: <http://doi.wiley.com/10.1111/j.1468-0262.2005.00594.x>.
66. DiPrete TA, Gangl M. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociol Methodol.* 2004; 34:271–310. Available from: <https://doi.org/10.1111/j.0081-1750.2004.00154.x>.
67. Abman R, Carney C. Land rights, agricultural productivity, and deforestation. *Food Policy.* 2020; 94:101841. Available from: <https://doi.org/10.1016/j.foodpol.2020.101841>.
68. Mahajan K. Back to the plough: Women managers and farm productivity in India. *World Dev.* 2019 Dec 1; 124:104633. Available from: <https://doi.org/10.1016/j.worlddev.2019.104633>.
69. Wuepper D, Roleff N, Finger R. Does it matter who advises farmers? Pest management choices with public and private extension. *Food Policy.* 2021 Feb 1; 99:101995. Available from: <https://doi.org/10.1016/j.foodpol.2020.101995>.