Agricultura, Sociedad y Desarrollo

EVALUATION OF THE RURAL EXTENSION POLICY IN MEXICO, BASED ON LINEAR AND NON-LINEAR ECONOMETRIC MODELS

Paulino **Benigno-Cruz**[1], Humberto **Vaquera-Huerta***[1], Gustavo **Ramírez-Valverde**[1], Sergio **Pérez-Elizalde**[1], Hilda Victoria **Silva-Rojas**[2], Victor Manuel **Santos-Chávez**[3]

1Colegio de Postgraduados, Campus Montecillo. Texcoco, Estado de México, México. Carrertera México-Texcoco Km. 36.5. 56264. Posgrado en Socioeconomía, Estadística e Informática-Estadística.
[2]Colegio de Postgraduados, Campus Montecillo. Texcoco, Estado de México, México. Carrertera México-Texcoco Km. 36.5. 56230. Posgrado en Recursos Genéticos y Productividad.
[3]Universidad Autónoma Metropolitana. Unidad Xochimilco. Calz. del Hueso No. 1100, Edificio F, 3er. Piso. Col. Villa Quietud, Alcaldía Coyoacán. 04960.
*Corresponding author: hvaquera@colpos.mx

Citation: Benigno-Cruz P, Vaquera-Huerta H, Ramírez-Valverde G ,Pérez-Elizalde S, Silva-Rojas HV, Santos-Chávez VM. 2024. Evaluation of the rural extension policy in Mexico, based on linear and non-linear econometric models.
Agricultura, Sociedad y Desarrollo
https://doi.org/10.22231/asyd.v21i3.1621

ASyD 21(3): 369-388

Editor in Chief:
Dr. Benito Ramírez Valverde

Received: May 29, 2023.
Approved: August 4, 2023.

Estimated publication date:
June 18, 2024.

ABSTRACT

Rural extension public policies are considered globally as a defining factor to accelerate agricultural innovation and to influence the reduction of rural poverty. Therefore, evaluating their results becomes important. The objective of this study was to evaluate the improvement in the income of recipients of the rural extension policy in Mexico during 2014-205. Data disaggregated at the level of plot by groups of recipients and non-recipients were used to evaluate the results based on the explicative variable of income. A database with statistical population of 1,083 producers was used, from which 58.5% were recipients and the rest non-recipients. Two regression models were adjusted to compare results. The multiple linear regression model based on ordinary least squares was adjusted in a first moment and, then, another non-linear generalized additive model with Pareto type dependent variable; in addition, a Bayesian additive regression tree model was used to verify the effect of the policy on the income of the recipients. The results exhibit that the generalized additive model with Pareto distribution and an identity link function was the best model, according to Akaike's information criterion. In the adjusted models, it was shown that being a recipient of the rural extension policy has a positive effect on the producers' income. This implies that the policy evaluated improved the income among its recipients.

Keywords: agricultural innovation, public policy evaluation, regression models.

INTRODUCTION

The rural extension systems and the public policies associated to providing technical assistance constitute actions and strategies that include activities of information supply and consultancy which the farmers and other stakeholders request in agrifood systems (Christoplos, 2010). According to Blockeel *et al.* (2023), these types of actions are seen as priority to achieve global objectives of transition toward more inclusive, resilient and sustainable systems. This implies addressing a complex set of socioeconomic problems, such as rural poverty, inequality, volatility of agricultural prices, hunger, but also environmental problems, such as soil degradation, biodiversity loss, and climate change. Nowadays, the role that rural extension systems can play in the attainment of these objectives, as part of agricultural innovation systems, is widely recognized (Läpple and Hennessy, 2015; Tollier *et al.*, 2021; UN, 2022).

In Latin America, rural extension policies were quite broad in different countries during the decades of the 1960s and 70s. In most of the countries of this region, the State was the main provider of this type of services, and the extension policy in the region was based on a centralized public system.

However, since the 1980s and particularly in the 90s, there was a dismantling process of the national rural extension systems; one of the main causes, according to Preissing *et al.* (2014), was associated to the high operative cost and uncertain impacts/results of these policies. In general, dissatisfaction processes could be seen with the results obtained by rural extension policies. In addition, there was a perspective that rural extension services have a low impact on the resolution of farmers' problems. At the same time, this translated into the policy not being able to improve the living conditions of the producers.

Various authors have pointed out that in recent decades there has been a new impulse for rural extension actions in Mexico and Latin America (De Rosa and Bartolli, 2017; Klerkx *et al.*, 2016); in this sense, it can be seen that extension work has recovered centrality in government and research agendas, due to the importance that it has been given as policy to overcome rural poverty, inequality and food insecurity, to increase productivity and the income of small-scale producers in rural areas. In fact, Preissing *et al.* (2014) emphasize that there is a renovated interest of Latin American governments in placing issues of rural development, including rural extension work, in the political agendas of governments, which without a doubt leads to the examination of results from these actions within the scope of public policies, since historically there was an absence of monitoring and evaluation that would allow understanding the short, medium and long term effects of this type of public policies.

Within this context, this study sets out to evaluate the results of the rural extension policy in Mexico, through the comparison of econometric models to select those that allow evaluating the impact of the rural extension policy during the period of 2014-2015 in a more robust way. The hypothesis of the research establishes that there was no positive effect of the extension policy on the income of recipients. Data from the monitoring and evaluation system of the Extension and Productive Innovation Component of the Integral Program for Rural Development in Mexico 2015 were used. The data contain information about the results of recipients and non-recipients. Given this set of observations, econometric models were compared, to measure the impact of the policy using producers' income as explicative variable.

THEORETICAL FRAMEWORK

Public policy evaluation: econometric approach

Rural extension work in Mexico and globally has recovered notoriety both in the government sphere and in research agendas, due to their transcendence in addressing rural poverty, inequality and food insecurity, through the transference of technologies and knowledge to rural producers, with the aim of increasing productivity and income, and, therefore, to stimulate rural development (Knook *et al.*, 2018; Cawley *et al.*, 2019).

According to the Global Forum for Rural Advisory Services (GFRAS), rural extension, as an activity developed in public policies, includes three large strategies: a) transference of technologies and information, which at the same time includes activities of knowledge exchange about markets, inputs, climate, among others; b) advice related to business and organizational management, which includes advice to individual farmers and producers' organizations, about ways to enter markets, as well as easing business management and support for institutional development and social innovation in organizations; and c) easement and implementation in rural development and value chains, which constitutes the collaboration in intermediation and promotion of links between farmers and their organizations, with public and private sectors, facilitating the feedback of groups of farmers and businessmen, for example, for access to credit, agricultural insurance, payment for environmental services, among others (GFRAS, 2012).

This allows understanding that rural extension work has been seen as a policy instrument that plays a prevalent role in improving productivity in general in the farming sector, through the increase of producers' income (Läpple and Hennessy, 2015); this leads to consider the evaluation of this type of interventions as a fundamental element.

The evaluation of public policies seeks to measure the way in which a public program has impacted the problems that it was designed to address; it is a modality that contrasts the impact of an intervention, for example, with the "before and after" or compares the impact of such an intervention between a group that was subjected to the intervention and another group that wasn't (control group); that is, contrasting what happened with what could have happened if there had been no intervention (Parsons, 2007).

A dimension within public policy evaluation is the one related to models of impact evaluation, which refer to the preponderant use of statistical methods and experimental approaches. These approaches seek to apply the principles of experimentation to social problems; they are based on the use of techniques where a problem is studied before and after the public intervention. Impact evaluation uses techniques that have the purpose of analyzing a control group and an experimental group, to later compare them and to measure the effectiveness of a specific public policy, with regards to a situation where there was no intervention (Parsons, 2007).

In the case of rural extension work, it represents a policy instrument whose purpose is the attainment of objectives of public interest to promote rural development, through knowledge transference processes to farmers and other stakeholders in agrifood systems (Santos *et al.*, 2019).

In this sense, it is essential to evaluate the achievements of the rural extension policy on the basis of empirical evidence, based on reliable methodologies that allow having a level of confidence about the direct effects of such a public policy. The impact evaluation would allow having a systematic and objective assessment of the policy, with which to determine the relevance and attainment of their objectives, as well as the effect on the income of producers, given that this variable has been defined as one of the objectives of the rural extension policy in Mexico in recent years.

The evaluation of public policies, in light of this, is an exercise that allows knowing the level of performance of a public program and to explain it taking into consideration the public problem that gave rise to it. This study stems from an evaluation approach that is directed toward the application of econometric techniques to measure the performance and to have information about the effectiveness and the results, considering for this purpose the impact on the income of recipients of the rural extension public policy in Mexico.

It is an evaluation whose focus is based on econometric techniques and procedures. Measuring performance is expressed in a comparative analysis between two groups: a group of people who were recipients of the rural extension policy, and another group of non-recipients. The evaluation carried out in this study seeks to focus on results, referring to the final consequences of the program for the people that it is aiming to serve.

In fact, one of the expected results of the extension policy was to increase the income of recipient farmers benefitted by this intervention. In this sense, this study stems from a general approach on impact evaluation, which claims that "…the meaning of impact is the effect or the net effects of the program, after considering what would have happened if it had not been implemented. From the operative point of view, it means examining the impacts for the program's participants and the impacts for an equivalent population that has not participated in the program…" (Weiis, 2015:57).

Consequently, taking an evaluation approach of broad impact, it analyzes the effect on the income in two groups of people to evaluate the results of a specific policy based on econometric methods; specifically in regression models, which according to Woolrige (2015), are very useful at the time of evaluating the effects of a policy when there are non-experimental data to depend on.

Multiple linear regression model

The multiple linear regression model is the most commonly used vehicle for the empirical analysis of the economy, and the ordinary least squares method is the most widely used to estimate the parameters of the model. According to Stasinopoulos *et al.* (2017), this model represents a simple but effective instrument that has served the statistical community in the last century. With a response variable *y*, *r* co-variables and a sample size *n*, it can be expressed as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_r x_{ir} + \varepsilon_i \quad (1)$$

where β_0 is the intercept; β_1 is the parameter associated with x_{i1}; β_2 is the parameter associated with x_{i2}; β_3 is the parameter associated with x_{i3}; ε_i is the error term or disturbance that contains other factors different from x_1, x_2, x_3, ..., x_r which can affect y_i. In this type of regression model, y_i has a distribution:

$$yi \sim N\left(\mu_i, \sigma^2\right)$$

where $\mu_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}+\ldots+\beta_r x_{ir}$ for $i=1,2,3,..,n$. To obtain the estimations of ordinary least squares, the model of the equation (1) is rewritten in vectorial form: $y=X\beta+\varepsilon$. Where $y=(y_1,\ldots,y_n)^{\mathrm{T}}$ represents the response vector, and the case of X represents the design matrix $n\times p$ with $p=r+1$; this contains r columns or co-variables, plus a column of ones. To estimate the β coefficients through the ordinary least squares method the sum of the square residuals is minimized, that is, given n observations on variable y, and on variables x_1 and x_2, $\{(x_{i1},x_{i2},y_i): i=1,2,\ldots,n\}$, the estimations of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ were chosen simultaneously for:

$$\sum_{i=1}^{n}\left(y_1-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\right)^2$$

to be as small as possible. Making a generalization for estimations of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2,\ldots,\hat{\beta}_k$ the following expression would be used:

$$\sum_{i=1}^{n}\left(y_1-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}-\ldots-\hat{\beta}_k x_{ik}\right)^2$$

The problem of minimization is solved by obtaining the conditions of first order of the ordinary least squares as exposed next:

$$\begin{aligned}
&\sum_{i=1}^{n}\left(y_1-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}-\ldots-\hat{\beta}_k x_{ik}\right)\\
&\sum_{i=1}^{n}x_{i1}\left(y_1-\hat{\beta}_0-\hat{\beta}_1-\hat{\beta}_2 x_{i2}-\ldots-\hat{\beta}_k x_{ik}\right)=0\\
&\sum_{i=1}^{n}x_{i2}\left(y_1-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2-\ldots-\hat{\beta}_k x_{ik}\right)=0\\
&\ldots\\
&\sum_{i=1}^{n}x_{ik}\left(y_1-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}-\ldots-\hat{\beta}_k\right)=0
\end{aligned}$$

These equations are known as first-order ordinary least squares. These equations are obtained with the method of moments, under the assumption of $E(u)=0$ and $\mathrm{E}(x_j u)=0$ where $j=1,2,\ldots,k$.

To simplify the expressions, the study resorts to matrix notation. The first-order conditions are simplified through the following formula:

$$X'\left(y-X\hat{\beta}\right)=0$$

$$\left(X'X\right)\hat{\beta}=X'y$$

Assuming that the symmetrical matrix $X'X$ of $(k+1)\times(k+1)$ is not singular, both sides of the expression $\left(X'X\right)\hat{\beta}=X'y$ can be multiplied by $\hat{\beta}$ of ordinary least squares and the estimator of ordinary least squares is:

$$\hat{\beta}=\left(X'X\right)^{-1}X'y$$

The generalized additive model for position, scale and shape

The generalized additive models for position, scale and shape (GAMLSS) constitute semi-parametric regression models. This type of models are, on the one hand, parametric, since they require starting from the assumption that the dependent variable has parametric distribution (beta, gamma or even normal type); and they are semi-parametric in the sense that modelling of the parameters of distribution is done independently, as functions of explicative variables using non-parametric functions (Stasinopoulos *et al.*, 2017).

The GAMLSS models, similar to the generalized linear models, assume that the dependent variable y follows a distribution of the exponential family with a mean that can be modelled in function of other variables (predictors) and whose variance is calculated through a constant of dispersion and a type $\nu(\mu)$ function.

This leads to consider, as specific element of the model, that the variance, asymmetry and kurtosis are not modelled indirectly in function of the predictor variables, through their relation with the mean.

The GAMLSS models assume independent observations y_i for i=1,2,...,n with a function of probability $f(y_i|\theta^i)$ conditioned to $\theta^i=(\theta_{1i},\theta_{2i},\theta_{3i},\theta_{4i})=(\mu, \sigma, \nu, \tau)$ a vector of four distribution parameters, each of them can be a function of explicative variables. In this model, the distribution parameters are μ, σ, ν, τ.

The original formulation, defined by Stasinopoulos *et al.* (2017), stems from the following premise: given $y^T=(y_1,y_2,...,y_n)$ known as the length vector n of the response variable, in addition to k=1,2,3,4, $g_k(.)$, known as the link functions that relate the distribution parameters to the explicative variables given by:

$$gk\left(\theta_k\right)=\eta_k=X_k\beta_k+\sum\nolimits_{j=1}^{jk}Z_{jk}\gamma_{jk} \quad (2)$$

where θ_k and η_k are length vectors $n,\theta_k^T=\left(\theta_{1k},\theta_{2k},...,\theta_{nk}\right)$, $\beta_k^T=\left(\beta_{1k},\beta_{2k},...,\beta_{J'_k k}\right)$ is a parameter of length J'_k, and at the same time X_k is a design matrix of order $n\times J_k$, and the variable Z_{jk}, represents a fixed design matrix $n\times q_{jk}$ and γ_{jk} is a random variable q_{jk}-dimensional.

The model of equation (2) is in fact defined as the GAMLSS model:

$$gk\left(\theta_k\right)=\eta_k=X_k\beta_k+\sum_{j=1}^{jk}Z_{ik}\gamma_{jk}$$

The vectors γ_{jk} for j=1,2,...,j_k can be combined within a simple vector γ_k with a simple design matrix Z_k. The formulation of the GAMLSS model allows combinations of different types of random effects.
The terms used within the context of the GAMLSS models to refer to the parameters of localization, scale and shape are (μ, σ, ν, τ). In the case of the parameters μ_i, σ_i they refer to parameters of location and scale, while the parameters ν_i, τ_i are characterized by parameters of shape, for example, parameters of asymmetry and kurtosis (Stasinopoulos *et al.*, 2007). With:

$$Y \sim D(\mu,\sigma,\nu,\tau), \text{ with Y=X}^T \beta$$

$$g_1(\mu) = \eta_1 = X_1\beta_1 + \sum_{j=1}^{j_1} Z_{j1}\gamma_{j1},$$

$$g_2(\sigma) = \eta_2 = X_2\beta_2 + \sum_{j=1}^{j_2} Z_{j2}\gamma_{j2},$$

$$g_3(\nu) = \eta_3 = X_3\beta_3 + \sum_{j=1}^{j_3} Z_{j3}\gamma_{j3},$$

$$g_4(\nu) = \eta_4 = X_4\beta_4 + \sum_{j=1}^{j_4} Z_{j4}\gamma_{j4}$$

where $Y \sim D(\mu,\sigma,\nu,\tau)$ is the distribution of the dependent variable (can be fewer parameters); X contains the linear terms of the model and β are the linear coefficients. $f_i(x_i)$ are non-linear smooth functions of each predictor.
According to Stasinopoulos *et al.* (2017), the GAMLSS model is more general than the generalized linear models and allows modelling dependent variables from the exponential family, although, in addition, more broadly, they are models that are not limited to this type of families, that is, they allow incorporating distributions that are not from the exponential family and modelling explicitly each of its parameters in function of the predicting variables, using linear and non-linear functions.

METHODOLOGY

This section presents the characterization of the database (DB) used to estimate the regression models and, then, the procedure is detailed for the estimation of the parameters based on the regression models. The adjustment of the models and descriptive analysis was carried out in the RStudio® software version 4.1.2.

Database used

During the 2014-2017 period, the Food and Agriculture Organization of the United Nations (FAO) and the then Ministry of Agriculture, Livestock Production, Rural Development, Fishing and Food (*Secretaría de Agricultura, Ganadería, Desarrollo Rural,* *Pesca y Alimentación*, SAGARPA), established a monitoring and evaluation plan that, in the case of the year 2015, included recipients and non-recipients of the rural extension policy at the national level.

The DB 2015, with a size of 1,083 producers distributed in 10 states of the Mexican Republic, was used. In each state included in the DB, stratified sampling with state representation was applied, with level of confidence of 95% and an error of 10% (SAGARPA, 2017).

The analysis of the DB reflects that it is a population formulated mainly by men (79.3%) that constitute small-scale farmers with an average age of 59 years and education equivalent to incomplete secondary school (7.2 years of study). On average, the surface of the production units is 19 hectares (h), the productive assets which they count on (infrastructure, machinery and equipment, and means of transportation), are equivalent to $97,556.4 MXN (2015=100). In terms of the reported income, these reach on average $267,689.8 MXN (2015=100). The extension groups where these producers participated were made up on average by 26.6 producers (Table 1).

From the total of the sample in the DB, 41.45% of the population is non-recipient, since at the time of the survey application, with a probabilistic design, these producers reported not having received the corresponding support, although the questionnaire was applied, with which there is a group to compare the results. In this sense, having two groups in the population studied is considered, whose information derives from a random process in the selection of sampling units and allows performing a study with public policy approach directed at measuring the result of the rural extension policy through the measurement of change in an economic income variable.

The empirical verification of the results on the income of the recipients of the rural extension policy is achieved with a methodological element that characterizes the DB,

Table 1. Characteristics of the database used in the analysis.

Estado	n	Sex (%)		Age	Education	Size of the plot (h)	Productive assets (MXN $)	Income (MXN $)	Size of the extension group
		M	W						
Chiapas	114	91.2	8.8	58.2	4.3	17.0	921.1	64,374.0	22.8
Guanajuato	119	84.9	15.1	60.6	6.6	19.4	77,537.8	584,967.0	21.1
Estado de México	123	69.1	30.9	57.0	9.3	5.3	123,375.6	153,618.8	18.5
Michoacán	145	82.1	17.9	58.7	7.2	22.5	125,840.0	247,737.4	24.5
Oaxaca	120	74.2	25.8	60.6	6.5	19.6	55,389.2	472,334.3	34.0
Puebla	148	68.2	31.8	58.6	7.2	6.9	23,850.0	68,672.2	35.3
San Luis Potosí	75	70.7	29.3	56.5	7.2	26.8	68,636.0	140,681.8	27.6
Sinaloa	61	86.9	13.1	58.8	9.5	29.1	186,221.3	561,091.8	17.3
Tabasco	71	87.3	12.7	62.4	7.3	19.5	66,944.8	165,967.7	48.1
Zacatecas	107	86.0	14.0	59.6	7.9	36.8	294,045.4	307,179.4	17.4
Totals/Average	1,083	79.3	20.7	59.0	7.2	19.0	97,556.4	267,689.8	26.6

Source: prepared by the authors based on SAGARPA (2017).

and which allows ensuring that the control group is very similar to the treated group, because it was selected through an *ex post* randomized process, where the survey applied identified which individuals received the support and which didn't. Next, the most important characteristics of the control group and the treated group are presented, which allow establishing that both groups are very similar (Table 2).

The average age of the group of non-recipients and recipients, respectively, is similar; in the year 2015, both groups reported an average age of 59 years. When inquiring about the education variable, it can also be confirmed that there is a similar level of education in recipients, equivalent to 7.35 years studying in the school system in recipients, while the non-recipients reported slightly lower education of 6.94 years studied.

The size of the production unit also does not present important differences. While the recipients reported having on average one production unit of 18.51 h, the non-recipients declared having plots of 19.77 h (Table 2).

The variable level of assets presents slight differences between the groups. The recipients have a greater investment in the level of assets, equivalent to $116,381.87; for their part, non-recipients have a level of capitalization of $70,974.45. The level of income of producers also shows substantial differences between both groups. The recipients of the rural extension policy in 2015 reported having an annual income of $320,047.31, while the non-recipients declared having annual income of $177,526.36 (Table 2). The descriptive analysis of these characteristics shows there are important differences in the two income variables (level of assets and income) of the groups studied.

Analysis of the variables included in the models

Given that the objective of the research is to obtain a model that allows evaluating the effect of the extension policy on the income, keeping the other factors that affect the income constant, such as sex, education, age, plot size, and surface of the production unit; the next section presents an analysis of the relationship between the income and the variables that can affect it.

Table 2. Characteristics of recipients and non-recipients 2015.

Variable	Recipients	Non-recipients
Frequency	634	449
Average age	58.9	59.2
Average education	7.35	6.94
Size of the average plot (h)	18.51	19.77
Level of average assets (MXN $)	116,381.87	70,974.45
Average income (MXN $)	320,047.31	177,526.36

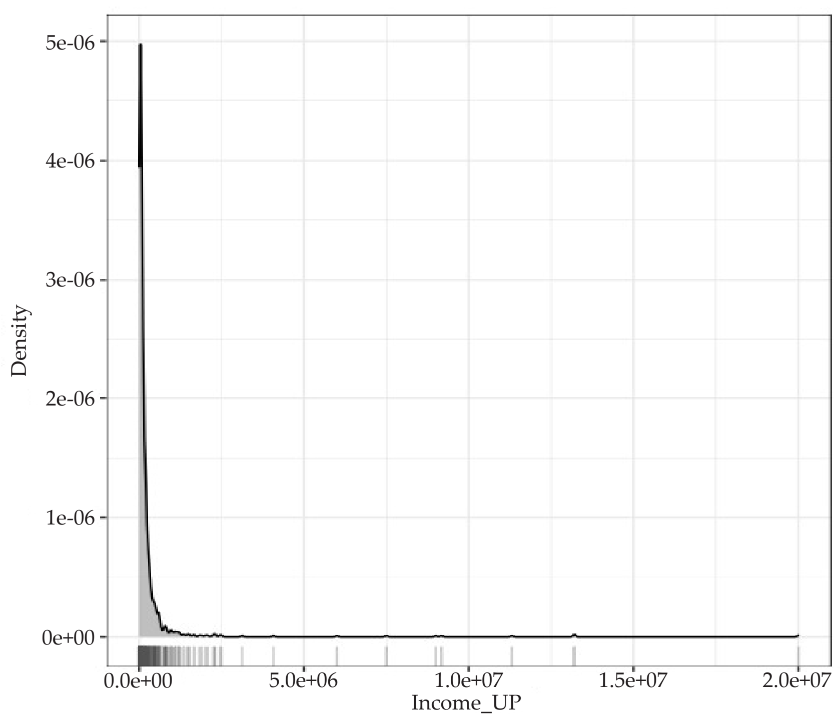
Source: prepared by the authors based on SAGARPA (2017).

Income of the recipients: Pareto distribution function

The density function of the producers' income (Figure 1) shows an asymmetrical curve, typical of the density functions of income; that is, a biased distribution with heavy tails. This type of distributions has been studied as Pareto-type distributions thanks to the contributions of the French economist Vilfredo Pareto, who observed that the number of people in a population whose income exceed *x*, very often approaches the expression $Cx^{-\infty}$.

At the international level, the analysis of different density functions of income have shown that a behavior similar to the expression established by Pareto can be expected only in the higher end of the income distributions, that is: $Cx^{-\infty}$ (Arnold, 2015). Consequently, Pareto's Law was established, which proposes that income distributions have heavy tails. Pareto type distributions, such as that of income (Figure 1), and their close relationships and generalizations, provide a very flexible family of distributions of heavy tails that can be used to model income distributions, as well as a broad variety of other social and economic distributions. This is why Pareto's Law and Pareto's distribution continue to be current themes (Arnold, 2015).

According to Arnold (2015), Pareto's distributions can be studied based on two models, the first expressed in the following way:



Source: prepared by the authors.

Figure 1. Probability density function of the income variable of producers.

$$\overline{F}(x) = Cx^{-\infty}$$

In this model, $\overline{F}(x)$ represents the proportion of individuals in a population where the income exceeds *x*, and therefore, for higher values or higher than *x* the study resorts to the expression $\overline{F}(x) = Cx^{-\infty}$. This expression is known as "Classical Pareto Distribution", from which different density functions stem, which are precisely extensions of the classic distribution. Pareto's classic distribution, with its respective biased distribution with heavy tails, is the most accepted model used to analyze income variables in modern theoretical economics. Based on this characteristic, modelling the income as dependent variable of the model to be used, will be based on a Pareto-type distribution; that is, it will adjust to a specialized regression model for Pareto-type income distributions. A Pareto II type distribution will be used, proposed by Arnold (2015), to model the income as explicative variable. For Pareto II type distributions, the model is given by:

$$f\left(x;\sigma;\kappa\right) = \frac{1}{\sigma}\left(1 - \frac{kx}{\sigma}\right)^{\frac{1-k}{k}} I\left(x > 0, \left(kx\right)/\sigma < 1\right) \tag{3}$$

where σ>0 and $-\infty < \infty$

In this model of density distribution, k=0 is obtained through the estimation of the limit $k \uparrow 0$. The density is type k<0. It should be noted that this characteristic results in an exponential density, which at the same time, corresponds to a distribution of the Beta family.

Once the type of dependent variable that will be used in the modelling is analyzed, then the relationship between the income variable and the independent variables is described.

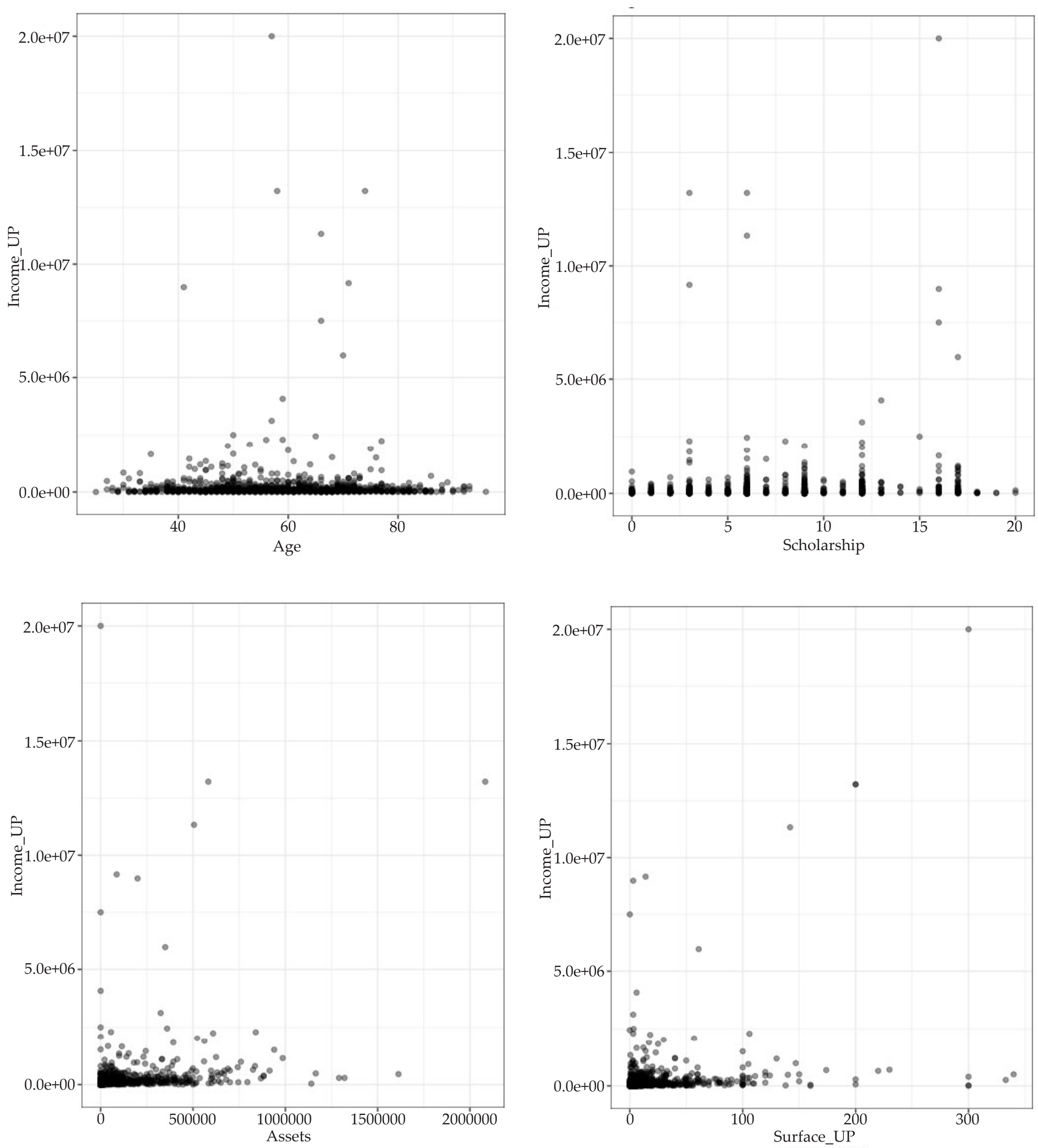
It can be seen that the relationship between income and variables education, age, level of productive assets, and size of the plot, is a positive relationship and seemingly not linear (Figure 2). To prove that the relationship is positive in every case, a Pearson's correlation analysis was used.

When verifying that the relationship between income and the other variables is positive, it can be assumed that as the age, education, size of the plot and assets of the production unit increase, the income of the producers increases (Table 3).

The analysis presented in the previous sections provides information regarding what effects to expect and what variables would be important to estimate differences between recipients and non-recipients of the rural extension policy. Next, the methodological elements of the regression models are presented.

Linear regression

Using the ordinary least squares method, the coefficients β of the following model were estimated:



Source: prepared by the authors.
Figure 2. Relationship of the income variable with the rest of the variables.

Table 3. Correlation analysis between the income variable and the rest of the variables.

Factor	Pearson's correlation	t	Df	p-value
Income and age	0.008461506	0.26705	996	0.7895
Income and education	0.09964073	3.1619	997	0.001615
Income and assets	0.2580386	8.4332	997	< 2.2e-16
Income and surface of PU	0.3246763	10.752	981	< 2.2e-16

Source: prepared by the authors.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

where y: income of producers; x_1: received the support or not; x_2: age of the producers; x_3: sex of the producers; x_4: education of the producers; x_5: size of the plot; x_6: productive assets of the plot; x_7: size of the rural extension group in which the producers participated.

The income of producers was used as dependent variable; a hypothesis about the distribution of this variable is that it is non-linear. This assumes that there could be a linear relationship between the dependent variable and the independent variables.
Next, the description of the dependent and independent variables included in the multiple linear regression model is presented (Table 4).

Table 4. Variables included in the multiple linear regression model.

Variable	Definition	Value and measurement unit
Dependent:		
Income of producers	Income of producers	Continuous (pesos from 2015)
Independent:		
Received the support or not	Producers who received the support or not in 2015	*Dummy* 0=Did not receive, 1=Did receive
Age	Age of the producers	Continuous (years)
Sex	Sex of the producers	*Dummy* 0=Man, 1=Woman
Education	Education of the producers	Continuous (years)
Size of the plot	Surface of the plot	Continuous (hectares)
Assets of the production unit	Assets of the plot in infrastructure, machinery and equipment, and means of transportation	Continuous (pesos from 2015)
Size of the extension group	Number of people who integrated the extension groups	Continuous (number of people)

Source: prepared by the authors.

GAMLSS model with Pareto-type dependent variable

In the adjustment of the GAMLSS model, the same dependent and independent variables were used than those used in the multiple linear regression model. Next, the formula of the probability density function is presented, which became operational in the RStudio® software for the corresponding regression model. The dependent variable was modelled as a Pareto II type variable, with a probability density function of the form:

$$f_Y\left(y \mid \mu,\sigma\right)=\frac{\sigma^{-1}\mu^{1/\sigma}}{\left(y+\mu\right)^{(\sigma+1)+1}}$$

For $y>0$, $\mu>0$ and $\sigma>0$.

Metrics to compare the models of each estimation method

In this section, the method to select the best model according to Akaike's information criterion (AIC) is presented. According to Faraway (2015), the AIC is used in practice, based on the Kullback-Leibler distance, which measures the approximation of the model calculated with real data; it is used to evaluate models, organized hierarchically according to the AIC value, and the one that minimizes this value is considered to be the best model. AIC is defined as:

$$AIC=-2\log\left(\mathcal{L}\left(\hat{\theta}\right)\right)+2K$$

where $\log\left(\mathcal{L}\left(\hat{\theta}\right)\right)$ is the maximum likelihood logarithm, which allows determining the values of the free parameters of a statistical model. K is the number of free parameters from the model.

Bayesian additive regression trees (BART) model

A Bayesian additive regression trees model (BART) was adjusted to test the hypothesis of the research. According to Hill *et al.* (2020), BART models provide a flexible approach to adjust a broad variety of regression models, while they avoid strong parametric assumptions. BART are integrated into an inferential Bayesian framework, to support the quantification of uncertainty and to provide an approach based on principles for regularization through prior specification. BART models are a model of the sum of trees to approach an unknown function *F*. The model handles over-adjustment in a way that is based on data, providing coherent uncertainty intervals.

Similar to other ensemble methods, each tree acts as a weak apprentice, explaining only part of the result. All these trees are of a particular type called decision trees. The decision tree is a very interpretable and flexible model, but it is also prone to over-adjustment. To avoid the over-adjustment, BART uses a prior regularization that forces each tree to be able to explain only one limited subset of the relationships between the co-variables and the predicting variables.

For a dimensional vector p of predictors x_i and a response $Y_i(1 \leq i \leq n)$ the BART model suggests:

$$Y_i = f\left(x_i, z\right) + \varepsilon_i,\ \varepsilon_{i\sim}^{iid} N\left(0, \sigma^2\right)$$
$$f\left(x, z\right) = g\left(x, z, T_1, M_1\right) + g\left(x, z, T_2, M_2\right) + \ldots + g\left(x, z, T_m, M_m\right)$$

where f is the sum of diverse regression trees. The construction block of BART is the regression tree. A regression tree creates a partition of the space of co-variables in subsets; the tree's adjustment will be the same for each observation in a given subset. In this equation, each of the functions g represents the adjustment of an individual tree, T_h represents the structure of the h^{th} trees, and the corresponding $Mh=(\mu h_1, \mu h_2, \ldots, \mu h_{bh})$ represents the group of subsets corresponding to the terminal nodes bh of the h tree.

RESULTS AND DISCUSSION

Estimations of the linear regression model's parameters

The results show that there are five statistically significant variables to explain the income variable. These are the intercept, the productive assets, the size of the plot, the producers' education, and the most important variable of the study: being a recipient or not of the rural extension policy. The goodness of fit analysis of the model, using an adjusted R^2, shows that 7 independent variables explain only 14.2% (R^2=0.1422) of the income variable variation. This slightly low value does not necessarily mean that the resulting equation of the model is not useful. The estimations of ordinary least squares, in this model, are useful to see the *ceteris paribus* effect on the income, being a recipient or not of the rural extension policy. The model shows that being a recipient of the rural extension policy has a positive effect on the recipients' income, compared to non-recipient producers.

This model shows that all the β estimators have a positive effect in the income variable of the recipients. Table 5 shows that the recipient producers of the rural extension groups have a higher income than those producers who do not participate in the extension actions in the year 2015.

The value of the AIC model was 27878 with a *p-value* of <2.2e-16. The following equation was obtained based on the multiple linear regression:

$$\widehat{y} = -370900 + 135200x_1 + 2111x_2 + 58820x_3 + 19980x_4 +$$
$$867x_5 + 0.9098x_6 + 538.6x_7$$

These results are in line with those found by Läpple *et al.* (2015 and 2013), who identified positive return rates in the income of producers of cow milk, who participated in rural extension programs in Ireland, compared to producers who did not participate in these policy actions.

It can be seen that there are two variables with high statistical level of significance to explain the income of producers, and they are: size of the plot and level of assets of the

Table 5. Estimation of parameters using multiple linear regression.

Variable	Estimation	Standard error	Value of t	Pr(>|t|)
Intercept	-370,900	211,900	-1.750	0.0804 .
Received backing: Yes	13,5200	73,230	1.846	0.0652 .
Age	2,111	2,883	0.732	0.4643
Sex: Woman	58,820	87,970	0.669	0.5039
Education	19,980	8,815	2.267	0.0237 *
Size of the plot	8671	988.9	8.768	< 2e-16 ***
Productive assets	0.9098	0.1854	4.906	1.1e-06 ***
Size of the group	538.6	669.2	0.805	0.4211

Level of significance: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ' 1; *p-value*: 2.2e-16; total observations: 911. Due to incomplete information, 172 observations were excluded.
Source: prepared by the authors.

production unit. It is congruent that the size of the plot is positively related with the level of income, and it is also a variable of income, since with larger expanse in production surface of a producer, higher production volumes can be obtained, and consequently, higher income.

This finding reveals congruence in terms of the economic theory in the signs of the estimators of the adjusted model. In the case of the variable productive assets, which includes infrastructure, equipment and machinery, and means of transport, also represents an income variable. Therefore, it is closely related to the income variable of the recipients. Thus, together with a surface variable, they are the two most important variables to explain the dependent variable.

Another outstanding finding which is in agreement with different studies that have analyzed the relationship between income and education, is the one related precisely to the relation in these variables in the adjusted model.

In the case of the rural extension policy, education is positively related to the income of producers, that is, as the school level of a recipient increases, there is positive probability of having higher income. The variables sex and age are not statistically significant at the time of explaining the behavior of the recipients' income.

Estimations of the GAMLSS Model parameters

The estimations of the GAMLSS Model parameters show an important difference with regards to the multiple linear regression model. Firstly, the sex parameter of the recipients changes the sign in the GAMLSS model, while the rest of the predictors of the parameters kept their sign (Table 6).

Based on the GAMLSS type regression, the following equation was obtained:

$$\hat{y} = 10.95 + 0.1735x_1 + 0.003807x_2 - 0.4498x_3 + 0.5995x_4 +$$
$$0.00629x_5 + 0.0000025x_6 + 0.000537x_7$$

Table 6. Estimation of parameters using GAMLSS type regression.

Variable	Estimation	Standard error	Value of t	Pr(>|t|)
Intercept	10.95	0.2836	38.616	< 2e-16 ***
Received backing: Yes	0.1735	0.09811	1.768	0.077378 .
Age	0.003807	0.003853	0.988	0.323389
Sex: Woman	-0.4498	0.1174	-3.833	0.000135 ***
Education	0.05995	0.01166	5.142	3.34e-07 ***
Size of the plot	0.006292	0.001115	5.642	2.25e-08 ***
Productive assets	0.000002507	0.0000001879	13.339	< 2e-16 ***
Size of the group	0.0005376	0.0008949	0.601	0.548154

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Sigma link function: log. Total observations: 911. Due to incomplete information, 172 observations were excluded. Degrees of freedom: 9.
Source: prepared by the authors.

Using the AIC value criterion, it can be determined that the GAMLSS model has a better performance compared to the linear regression model. The value of AIC in least squares is 27878 and in GAMLSS it is 23502.

It is important to highlight that the analysis of the most important explicative variable for the purpose of the study shows that, in general, being a recipient of the extension policy has a positive effect on the income of recipient producers, compared with the group of non-recipients of this policy. This leads to establishing that the actions implemented by the Rural Extension Component in Mexico during 2015 helped to improve the income of its recipients. The positive effect in the income, derived from the estimation of the coefficient of the explicative variable "Received backing: Yes", shows a result in the medium term, since the result from the intervention in the 2014-2015 period is being evaluated. It is necessary to point out that the probability of the positive effects continuing in the long term on recipients could change if the exercise is conducted for a longer period. However, the analysis of the circumstantial empirical evidence from the period studied showed a positive result in the income of recipients.

The results from the research correspond with those that were found by Läpple and Hennessy (2015), who detected higher income and productive yields between the recipients of extension groups in Ireland, specifically, in the case of dairy cattle farming. In addition, Lyne *et al.* (2018) also found that participating in extension groups contributes to increasing the income of small-scale farmers in a region of South Africa. In both cases, it is interesting to mention that they are rural extension actions designed as public policy, and that extension work is seen as a public good; in addition, the results are also based on econometric methods.

An interesting result that contrasts in the two models is relative to the sex variable. Meanwhile, in the linear regression model, this variable was not statistically significant and with an estimation in the parameter calculated positively, in the GAMLSS model, it is a negative beta coefficient and statistically significant (*p=0.000135 ****).

This is especially useful in terms of the innovation methods used in the extension policy, since the results show that different extension actions carried out have a differentiated effect between men and women; in general, the evidence shows that if a woman participates in the extension groups there is a high probability of her income decreasing, compared to men. This could be a problem that requires careful reflection, in terms of the challenges of the policies at the international level, regarding the importance of the gender approach in agriculture. This finding agrees with the results by Santos *et al.* (2023), who found that women in Mexico who participate in the innovation groups derived from the rural extension policy have lower probability of adopting practices and technologies in their plots.

Analysis of the Bayesian Additive Regression Trees model

The BART model is computationally efficient, compared to its competitors, including last-generation implementations for continuous, binary and categorical results. In this case, the response variable was the income of the production unit, and the treatment was converted into categorical variable in 0 for those who did not receive backing and 1 for those who did receive backing. The confusion factors were the co-variables of age, education, size of the plot, surface of the production unit, and productive assets.

A collection of treatment and response models was adjusted, using the Bayesian Additive Regression Trees (BART) algorithm, which produces estimations of the effects of the treatment through the BART function in the RStudio® software.

The estimations are adjusted starting from 911 total observations, with a reliable interval of 95%, calculated by: normal approximation of the TE population approached by: posterior predictive distribution. This result is based on 100 posterior samples multiplied by 10 chains.

According to Table 7, with the inferior and superior interval of confidence of 5.908 at 28.1, the contention of 0 in this interval is rejected, so the conclusion is that there is an effect from being a recipient on the production unit's income, which reinforces the results obtained with the previous models studied.

CONCLUSIONS

In terms of public policy, specifically of public policy evaluation, the regression models were useful to confirm that there is a positive effect in the income of the recipients of the extension policy, compared versus a control group of producers who are non-recipients

Table 7. Result from the Bayesian Additive Regression Trees model.

	Estimation	Standard deviation	Lower interval of confidence	Higher interval of confidence
ATE	17.01	5.665	5.908	28.11

Source: prepared by the authors.

of the policy. In both models, the estimation of the beta coefficient corresponding to the binary variable (Received backing: Yes/No) is positive for the producers that did receive the support ("Yes"). This suggests that the rural extension policy can play an important role in the offer of services directed toward small-scale farmers and rural populations, with the objective of increasing their knowledge and improving their income and thus contributing to the general welfare in the rural sphere.

The multiple linear regression model is useful to demonstrate that there is a positive effect in the incomes of recipient producers; however, the model presented R^2 value of 0.1422, which forces to consider a more robust model to estimate the effect of a public policy on the income of its recipients.

The GAMLSS type model is more robust than the classic linear, at least to model the effect that a public policy has on an income variable with Pareto II distribution. This model presented the lowest value in the AIC, compared against the respective model of Bayesian Additive Regression Trees; the conclusion is that there is an effect from being recipient on the producers' income, which reinforces the results obtained with the prior models studied.

The use of these models represents a relevant example based on the use of empirical evidence for the analysis and evaluation of public policies, where there is interest in knowing the effect of public programs on an income variable, in this case, the effect of a social program of rural extension on the income of the recipient producers.

REFERENCIAS

Arnold B. 2015. Pareto Distributions. Monographs on Statistics and Applied Probability 140. Second Edition. Taylor & Francis Group. A Chapaman & Hall Book. United States of America, 2015. 408 p.

Blockeel J, Chuluunbaatar D, Holley A, Sulaiman R, Djamen P, and Grovermann C. 2023. Taking a snapshot of Extension and Advisory Systems performance and ocutcomes: insights on a semi-quantitative evaluation approach. The Journal of Agricultural Education and Extension. 29(4). 489-509. DOI: 10.1080/1389224X.2022.2089178.

Cawley A, O'Donoghue C, Heanue K, Hilliard R, Sheenan M. 2019. The impact of agricultural knowledge transfer resources on farm level profitability during the economic recession – a quantitative study. The Journal of Agricultural Education and Extension. 25(3). 181-177. DOI: 10.1080/1389224X.2019.1583816

Christoplos I. 2010. Cómo movilizar el potencial de la extensión agraria y rural. 1ª Edición. Organización de las Naciones Unidas para la Agricultura y la Alimentación. Roma, 2010. 68 p.

De Rosa M, Bartoli L. 2017. Do farm advisory services improve adoption of rural development policies? An empirical analysis in GI areas. The Journal of Agricultural Education and Extension. 23(5). 461-474. DOI: 10.1080/1389224X.2017.1347099.

Faraway J. 2015. Linear models with R. Second Edition. Editorial: Taylor & Francis Group. United States of America, 2015. 270 pp.

Foro Global para los Servicios de Asesoría Rural (GFRAS). 2012. Guía para evaluar la extensión rural. 1ª Edición. Foro Global para los Servicios de Asesoría Rural. Suiza, 2012. 60 p.

Hill J, Linero A, Murray J. 2020. Bayesian Additive Regression Trees: A Review and Look Forward. Annual Review of Statistics and Its Application. 7(1). 251-278. DOI: 10.1146/annurev-statistics-031219-041110

Klerkx L, Landini F, Santoyo-Cortés H. 2016. Agricultural extension in Latin America: current dynamics of pluralistic advisory systems in heterogeneous contexts. The Journal of Agricultural Education and Extension. 22(5). 389-397. DOI. 10.1080/1389224X.2016.1227044.

Knook J, Eory V, Brander M, Moran D. 2018. Evaluation of farmer participatory extension programmes. The Journal of Agricultural Education and Extension. 24(4). 309-325. DOI:10.1080/1389224X.2018.1466717.

Läpple D, Hennessy T. 2015. Assessing the impact of financial incentives in extension programmes: Evidence from Ireland. Journal of Agricultural Economics. 66(3). 781-795. DOI: 10.1111/1477-9552.12108.

Läpple D, Hennessy T, Newman C. 2013. Quantifying the economic return to participatory extension programmes in Ireland: An endogenous switching regression analysis. Journal of Agricultural Economics. 64(2013). 467–482.

Lyne M, Jonas N, Ortmann G. 2018. A quantitative assessment o fan outsourced agricultural extension service in the Umzimkhulu District of KwaZulu-Natal, South Agrica. The Journal of Agricultural Education and Extension. 24(1). 51-64. DOI: 10.1080/1389224X.2017.1387159.

Parsons W. 2007. Políticas Públicas. Una introducción a la teoría y la práctica del análisis de políticas públicas. 1ª Edición. Facultad Latinoamericana de Ciencias Sociales/Miño y Dávila Editores. México, 2007. 816 p.

Preissing J, Ardilla S, Buitrón J, Fernández C. 2014. Nuevas inversiones en extensión para la agricultura familiar. *In:* Agricultura familiar en América Latina y el Caribe: Recomendaciones de política. Primera Edición. Salcedo, S. y Guzmán, L. (eds). Organización de las Naciones Unidas para la Alimentación y la Agricultura. Chile. pp: 215-231.

Santos V, Álvarez A, Pérez F, Pérez L. 2019. El extensionismo rural mexicano: análisis coyuntural con enfoque de políticas públicas. Revista Mexicana de Ciencias Agrícolas. 10(1). 63-77. DOI: https://doi.org/10.29312/remexca.v10i1.1117.

Santos V, Arana O, Martínez M, Garza L, Mora J, Santoyo H. 2023. Determinantes de la innovación agrícola entre pequeños agricultores en México. Una aplicación econométrica del Modelo Tobit. Agrociencia. 57(2). 339-368. DOI: https://doi.org/10.47163/agrociencia.v57i2.2878.

SAGARPA (Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación). 2017. Compendio de indicadores de gestión y resultados 2015. Componente de Extensión e Innovación Productiva del Programa Integral de Desarrollo Rural. 1ª Edición. SAGARPA. México. 85 pp.

Stasinopoulos M, Rigby R, Heller G, Voudouris V, De Bastiani F. 2017. Flexible Regression and Smoothing using GAMLSS in R. Taylor & Francis Group. A Chapaman & Hall Book. United States of America, 2017. 550 p.

Toillier A, Mathé S, Saley A, Faure G. 2021. How to assess agricultural innovation systems in a transformation perspective: a Delphi consensus study. The Journal of Agricultural Education and Extension. 28(2). 163-185. DOI: 10.1080/1389224X.2021.1953548.

United Nations (UN). 2022. The Sustainable Development Goals Report 2022. United Nations Publications. United States of America, 2022. 68 p.

Weiss C. 2015. Preparando el terreno. *In:* Antología sobre evaluación. La construcción de una disciplina. Primera Edición. Maldonado, C. y Pérez, C. (Coords.). Centro de Investigación y Docencia Económicas, A. C. México. pp: 43-84.

Wooldridge J. 2015. Introducción a la econometría. Quinta Edición. CENGAGE Learning Editores S.A. de C.V. México. 878 p.