

Modelling Technology Adoption in a Social Interactions Framework: Evidence on Improved Cassava Varieties in Colombia



Jairo Eduardo Gómez Chamorro

Supervisor: Juan Esteban Carranza Romero

Ph.D in Economics
Department of Economy
Yale University

This Thesis is submitted for the degree of
Master in Applied Economics

October 2017

I would like to dedicate this thesis to my loving parents and siblings who have inspired me to continue reaching new goals in life.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Jairo Eduardo Gómez Chamorro

October 2017

Acknowledgements

I would like to acknowledge International Center for Tropical Agriculture (CIAT) and Ricardo Labarta for providing me with the opportunity of participating in the project through the data analysis and sharing with an outstanding multidisciplinary team which did the best to understand the most challenging issues found in small-scale peasant agriculture.

Abstract

This study analyzes the impact of social interactions on the adoption of a new technology. Based on the basic model of social interactions with non-cooperative choices, we provide empirical evidence of spillovers or neighborhood effects on the individual probability of adoption of improved cassava varieties in the State of Cauca, Colombia. We use a sample of socio-economic characteristics at the household level and a set of covariates at the municipal level for cassava small-holder farmers to test peer effects on individual choices. We implement a machine learning clustering algorithm that allows us to measure the information channel and the degree of interaction between adopters and non-adopters by using the household's geographical coordinates to calculate physical distance using the Global Position System (GPS). The regression results show that the average village adoption rate has a significant effect on the household's decision to use modern, improved varieties of cassava seeds. The results also suggest that, as the minimum distance in kilometers by road from the non-adopter to an adopter increases, the probability of adoption decreases. Our results are robust to the so-called correlated non-observables identification problem widely recognized in literature on social interactions models. Finally, this work leads to highlight the role of simultaneity or reflection problem also widely identified in theory. The empirical evidence does not support the presence of simultaneity. However, the future analysis requires detailed information at agent's network level. We also provide suggestive evidence of potential learning effects by conditioning on adopters who are technologies specific-users such as weed control, fertilizers, etc. This is in accord with stylized facts that highlight the importance of knowing more about techniques in the process of adoption. The results show no difference in the estimated coefficients between the neighborhood effects and the suggestive evidence of learning. Future research needs to build better identification mechanism as part of the analysis on social networks, diffusion and learning. This study represents a first step for understanding the role of peer effects on the improvement of living conditions for rural population via technological enhancements.

Key words: *Social interactions model*Peer effects*Cassava crop*K-means clustering algorithm*Logistic regression.

Table of contents

List of figures	vii
List of tables	viii
1 Introduction	1
2 Background and Methodology	6
2.1 Background on cassava	6
2.2 The Model	10
2.3 Econometric Approach	13
3 Empirical Analysis and Results	16
3.1 Data Set and Sample Description	16
3.2 Empirical Measures of Social Interactions	20
3.2.1 Closeness and Betweenness Measures	22
3.3 Regression Results	23
3.4 Discussion	28
4 Concluding remarks and future research	31
References	32
Appendix A Graphs and Tables	35

List of figures

2.1	World Production Index	7
2.2	World Area Harvest Index	8
2.3	World Area Yield Index	8
3.1	Dendrogram	18
A.1	Spatial distribution of adopter and non-adopter households	35
A.2	Spatial distribution of adopter and non-adopter households (driving mode) .	36
A.3	Clustering	37
A.4	Density	38

List of tables

A.1	Variables Description	39
A.2	Summary Statistics	40
A.3	Adoption Rate by Village	41
A.4	Regression Results I	42
A.5	Regression Results II	43
A.6	Regression Results III	44
A.7	Regression Results IV	45
A.8	Regression Results V	46
A.9	Regression Results VI	47
A.10	Logistic Regression Results	48

1. Introduction

Technology adoption plays a key role as a mechanism to reduce income and poverty gaps and to improve life conditions. This is common especially in developing countries where farming households face challenges associated with scarce access to new technologies, financial and credit support, limited access to information, insufficient human capital and regular agricultural practices to raise productivity in crops.

A significant body of research in agricultural economics is devoted to analyzing the problem of choosing between two distinct technologies: traditional or unimproved varieties of seeds and high-yielding varieties (HYV's)¹ (Feder et al., 1985). The theoretical and empirical approaches highlight *learning* as the main force that drives the adoption process, since farmers follow a Bayesian inference along the period in which they decide whether or not to adopt the new HYV consistent with yielding and seed performance (Besley and Case, 1993; Foster and Rosenzweig, 1995; Cameron, 1999). In other words, when a new technology is introduced, at first, agents involved in the process are unfamiliar with it and have subjective beliefs about outcome distribution shaped by prior experiences. The experimentation along the process enables agents to learn more about their own experiences, updating their posterior beliefs (Baerenklau, 2005). Nonetheless, this kind of approach entails a set of advantages and challenges in practice.

As Cameron (1999) and Baerenklau (2005) point out, the use of panel data allows to estimate the *learning by doing* term because the array of farmers' socio-economic characteristics can be traced over time. This leads to the most common measure of learning being computed as the average profit differential between the HYV and traditional seed. However, panel data is very uncommon to observe due to the hardness of keeping track of plots and households over a time period. Furthermore, the few existing surveys that employ panel data can not account for the other overriding force to conduct to adoption process as well: *learning by neighborhood*.

¹We follow the microeconomic approach based on adoption at agent's choice level. For more detailed information about adoption at aggregate level and the so-called adoption S-curve, see (Feder et al., 1985; Griliches, 1957) respectively.

Due to the scarce availability of panel data, a number of studies have analyzed the adoption of new agricultural techniques by using a cross-sectional sample of socio-economic characteristics at the household level. This approach, analyzed in the vast majority of applied exercise, focuses only on showing correlations amid the variables involved across the sample without conducting to any specific motivation behind the actual process of interest. In other words, the analysis only reduces to show a regression of an adoption measure against farmers' socio-economic characteristics (Acheampong and Owusu (2015), Afolami et al. (2015), Zeddies et al. (2001)) . Albeit seemingly cross-sectional samples may be unable to provide an insightful view about the way of approaching to learning measures, is not always the case. As a matter of fact, the real problem lies in disentangling proxies for learning in a social environment. Finding feasible empirical approaches require new measures of social networks and geographical patterns that deal with the lack of information not included in traditional analysis (Foster and Rosenzweig, 1995; Cameron, 1999). Cameron (1999) warns that approaching to learning by neighborhood can be much more difficult than learning by doing because the observance of individual's social networks are not directly available in data. However, geographical patterns may be easier to collect and works as a mean of communication and diffusion by which agents interact each other, sharing information about new agricultural practices, among other things. For instance, a farmer who does not use new improved seeds of a crop is more likely to adopt them if his nearest neighbor does use them. To some extent, this degree of social interaction sets a scenario in which agents learn and exchange information each other, allowing to measure a learning channel between farmers. Even so, learning by neighborhood captured through physical distance is not itself a guarantee that the learning process can be carried out. On this, some household farmers may prevent from learning from other farmers whether they face technological restrictions as we will analyze later on.

The interest in analyzing and measuring social interactions has been widely recognized in the theoretical literature². Banerjee (1992) formalizes the agents' decision making under herding behavior, demonstrating that an optimal decision faced by an agent relies on what everyone else is doing. Akerlof (1997) develops a model of social distance where agents nearest each other interact strongly than those who have little closeness. In general, the underlying force behind those intuitions that contributes to influence individual actions through social behavior is defined as *neighborhood* or *peer effects*. In other words, peer effects capture the idea that agent's behavior can be affected or influenced by exposure to the behavior of other agents.

²See Manski (1993, 2000) for further details on social analysis from sociologist and economic point of view.

A branch of research in microeconomics discusses the correlation between individual and group decisions in a wide extension of topics. For instance, Coleman et al. (1966) emphasize that educational underachievement in disadvantaged students would ameliorate by exploiting the link between peer group reliances rather than incurring in additional school costs. Bertrand et al. (2000) measure the impact of contact availability in the same spoken-language group within a neighborhood which a person belongs to on the chance of being admitted in the welfare system of an specific country. Conley and Topa (2002) analyze unemployment and search job clustering through several social and economic distance metrics like physical distance and differences in ethnic distributions between locations. Borjas (1995) exposes that the mean skills of the ethnic group in the parent's generation affect the skills of the ethnic children, apart from parental skills. Similarly, a sizeable impact of network effects are present in literature related to crime (Glaeser et al., 1996), female labor supply (Woittiez and Kapteyn, 1998), diffusion and technology adoption in computers (Goolsbee and Klenow, 2002) and electronic payment system (Gowrisankaran and Stavins, 2004), cultural beliefs (Marini, 2011) and adoption of new agricultural techniques (Bandiera and Rasul, 2002; Conley and Udry, 2010; Yishay and Mobarak, 2015; Beaman et al., 2015).

In the same way, macroeconomists have recently stressed the meaningful role of diffusion and learning process for long-term economic growth by taking *spillover* into account as indicated by the endogenous growth theory (Romer, 1986; Lucas, 1988). Most recently, new surveys on aggregate fluctuations shed light on how intersectoral input-output linkages determine the decay rate of aggregate volatility regarding the specific hubs and cascade effects in the network structure (Acemoglu et al., 2012; Carvalho and Voigtländer, 2014).

This document aims at providing strong empirical evidence for social interactions model based on the basic setup of the Brock and Durlauf (2001) model, in the adoption of improved cassava varieties in the State of Cauca, Colombia. By using a unique data set of 293 farming households' socioeconomic-characteristics of a survey carried out by the *International Center for Tropical Agriculture* (CIAT) from November 2015 until January 2016 and a rich set of covariates provided by the *Sistema De Información Geográfica Para La Planeación Y El Ordenamiento Territorial* (SIG-OT), we find that: the household's average collective adoption in a group which belongs to, significantly determines the individual propensity to adopt. In order to provide more evidence on the prior outcome, we exploit the household's geographical coordinates to measure the information channel between adopters and non-adopters. Through the use of a machine learning clustering algorithm, our analysis suggests that as the minimum distance between a non-adopter and an adopter increases, the probability for the non-adopter decreases. The evidence found captures how the influence in terms of

closeness of the nearest adopter neighbor affects the non-adopter to adopt. In this sense, we identify the social effect of the learning by neighborhood by means of how close the agents are each other. This effect works as information sharing between the sides, considering as a stylized fact that non-adopter farmers would be willing to adopt new improved seeds if a neighbor would do the same. On the other hand, farmers face restrictions regarding the agronomic variety management. One reason that makes farmers less likely to adopt is the lack of knowledge about how to tackle the technological management of the new improved seeds. Nothing guarantees that living near an adopter makes a non-adopter become adopter. Maybe if an adopter who knows more about agricultural techniques can force a non-adopter into becoming adopter. We also identify this effect by conditioning if an adopter is also a farmer trained in the use of extensive services. The previous one represents the most striking result as a way to assess the interaction and spillover effects found in literature. Toward this end, we also compute closeness and betweenness centrality descriptive measures extricated from the social network of households to highlight the role of diffusion on the probability of adoption.

The contribution of our approach consists in providing new empirical evidence on social interactions models for the case of cassava crop in Colombia. First of all, the empirical implementation of those models has been difficult to apply due to the detailed information required in practice. We fill this gap by also providing a new measure of neighborhood effects in order to support the main prediction derived of this kind of models. Secondly, the lack of knowledge on cassava crop has created a growing interest in this field over the past few years, promoting new lines of research in multidisciplinary approaches. This work contributes to highlight the main findings from the agricultural economics perspective. Finally, this is the first time this approach is used in Colombia.

The literature on social interactions has raised concern over identification problems related to *correlated non-observables* and *simultaneity* or *reflection problem* (Manski, 2000; Brock and Durlauf, 2000; Bandiera and Rasul, 2002). Primarily, Bandiera and Rasul (2002) mention that the most prominent studies do not observe information at individual's networks level³ thus, those ones are forced to correspond to observed measures of geographical or cultural proximity. By doing so, villages⁴ averages are used as a proxy indicator by assuming the individual behavior is a function of the mean behavior of the remaining individuals belonging to the same village. The previous one implies that social effects can not be separated from unobservable characteristics at the village level. This is not a real concern in this work

³An exception is Beaman et. al (2015) who detail information at farmer's network level depending upon farmers targeted and trained on the new technology.

⁴Henceforth, we use the term village and municipal interchangeably.

because we have information at that level of detail, enabling us to control for unobservable characteristics. This point is, to some extent, crucial for our analysis as we can be closer to a reliable result in comparison with the previous studies in which those characteristics are omitted.

In respect of the second identification problem, Manski (1993) defines it as a feedback mechanism between individual and group behavior. In other words, the underlying idea behind simultaneity means that the group's behavior has an effect on the individual behavior which in turn affects group's behavior. From the empirical point of view, this concern is more complicated to explain due to inobservance of the agents' network structure. We attempt to reduce the consequences by using the social network distance among the households. However we are aware that we need better information based on individual network structure. Finally, the dynamic nature of the adoption choice entails restrictions due to the use of the cross-sectional sample and possible bias by the learning by doing term omission (Besley and Case, 1993; Cameron, 1999). This issue arises because advantages of adoption depend on the history of preceding adoptions accumulate up to time t (Bandiera and Rasul, 2002). We are not entirely subject to this problem since some of the new varieties bred across the farmers have recent history.

This document is organized as follows. In the next section we present a short background information and aggregate statistics on cassava crop. We also explain the methodology and the challenges related to theoretical and empirical approaches discussed in the previous literature. Section 3 goes through empirical measures construction and the regression results. Section 4 concludes.

2. Background and Methodology

2.1 Background on cassava

Manihot Sculenta Crantz, the scientific name as Cassava¹ is known, has played a relevant role over the last two decades. Its wide array of properties has made it become an optimal crop for full scale production among the countries belonging to the Western hemisphere tropical region, including Colombia. According to Howeler et al. (2013)², cassava is a root crop grown by almost 800 million people world-wide, mainly by low-income smallholder farmers. It is one of the richest and cheapest sources of energy based on high calories being consumed by many households under extreme poverty conditions, whereby it is called "the food of the poor".

Cassava leads the position as the 21st century crop not only for being a significant mechanism toward reducing hunger but also for overcoming and looking deeper into food security challenges, climate change and economic development. Going through those fields, respectively, cassava roots are both consumed by farming households and fed to their livestock, whose roots are processed for different uses such as paper and textiles production, fructose, alcohol and ethanol among others. Furthermore, due to its ease of producing on marginal areas with poor soil and unpredictable rainfalls, cassava does not require necessarily complex mechanization processes and inputs to raise productivity, allowing it to also be an adaptable crop for increasingly higher warm temperatures caused by climate change. For these reasons cassava is set to be a multi-purpose crop.

To place in context summary statistics on the behavior of cassava on some fronts, Figure 2.1 depicts a World Production Index of five main crops between 1980 and 2014. As illustrated in the figure, cassava ranks second after maize, surpassing other important crops like rice, wheat and potato. This measure is indicative of the growing importance that the crop

¹Cassava is also called Mandioca or Yuca in some Spanish-speaking countries such as Mexico and Colombia.

²FAO stands for Food and Agriculture Organization

has reached over the past few years in terms of production.

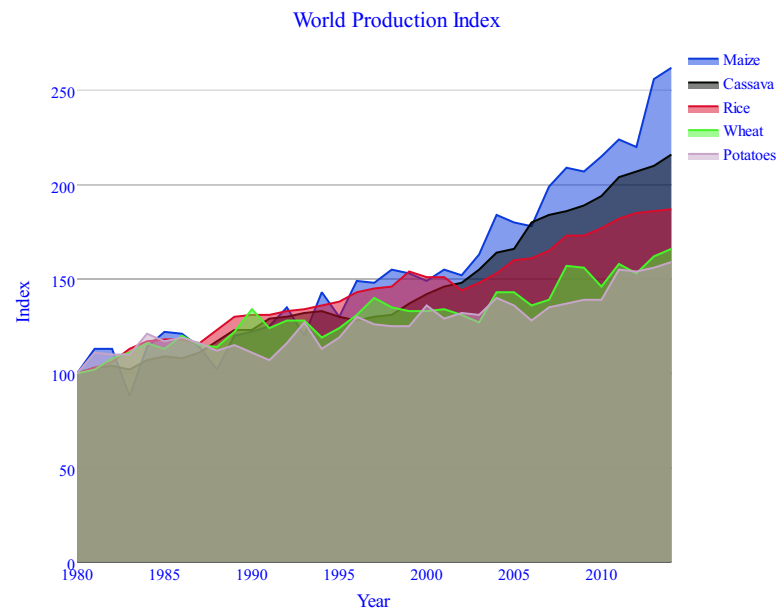


Fig. 2.1 Source: FAO. Author calculations.

Taking a look at the Figure 2.2, the World Area Harvested Index exhibits an indisputable leadership of cassava to the rest of the main crops cultivated world-wide, while the trend in rice, potato and wheat is toward a stationary level. Finally, Figure 2.3 plots a World Yield Index in which cassava is not well positioned in this area, positioning itself at the bottom of the ranking. A simple narrative of the facts set out on the previous charts indicates that, in the case of cassava, the long-run trend in production may be associated with increases in area sown rather than harvest yield. This is consistent, to some extent, with the growing interest in the rapid expansion of cassava production across a broad range of sectors and uses. It is also noteworthy to mention that the comparative poor performance of cassava yield is compatible with the limited research done so far on new improved varieties intended to improve and foster the conditions of high-yielding crop³.

³The development and genetic recognition of improved varieties intended to increase the crop yield is under investigation. For instance, our data suggest particular cases where a seed variety belongs to an only genetic family. This means that we ignore if a seed falls in the category of improved or unimproved variety. Perhaps the low degree of knowledge on the fact has slowed down the adoption of improved varieties that aim to increase yield

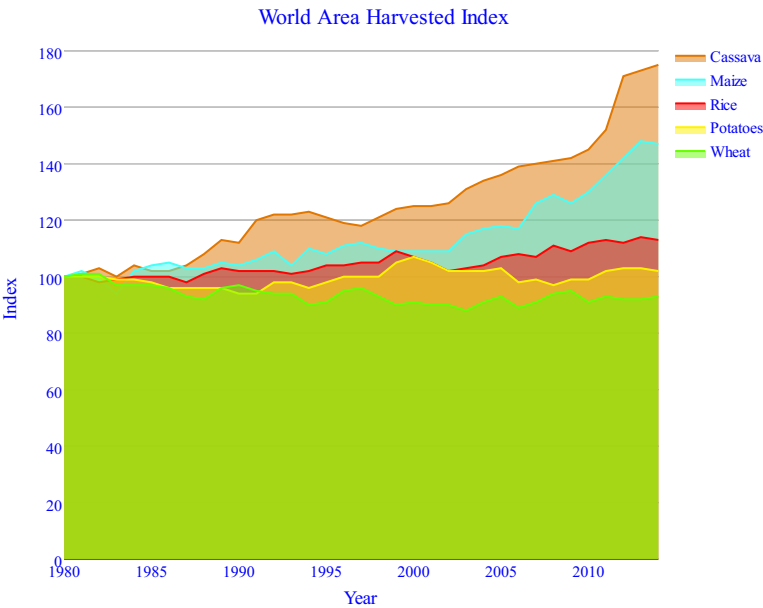


Fig. 2.2 Source: FAO. Author calculations.

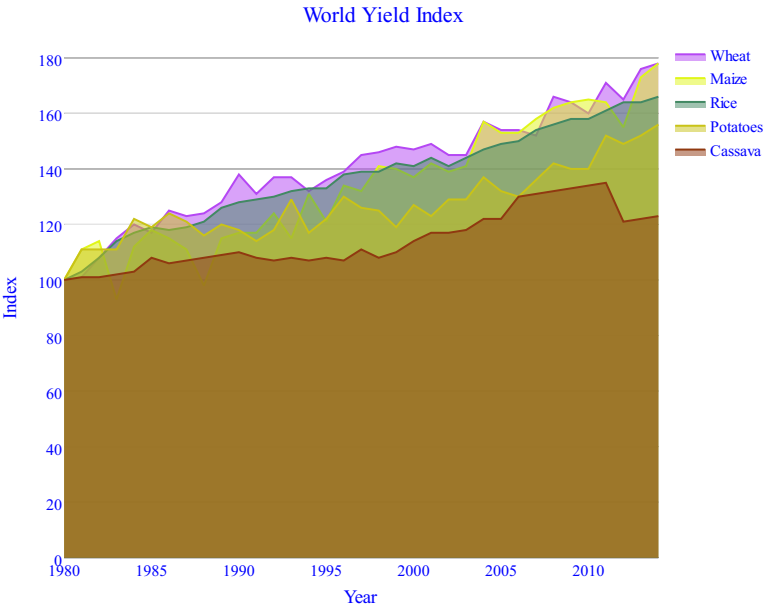


Fig. 2.3 Source: FAO. Author calculations.

In light of the above, whilst it is true that cassava promises to be an exceptional crop due to the wide set of advantages that derives, there are many challenges to deal with on the way yet. First of all, the planting cassava process carried out by farmers on degraded soils without previous tillage produces lower yields. In this sense, two practical ways of restoring the soil health state are: to include intercropping by mixing different types of crops in order to take advantage of mineral nutrients and protective hedgerows to reduce soil erosion. Nonetheless, untilled land under healthy conditions is able to produce higher yields at a lower cost. Secondly, to reach the full potential production of cassava and to face with certain constraints associated to good-yielding seeds, specific-agro-ecological practices, cropping systems and end-uses, the developed of improved varieties has an essential role for sustainable intensification. The adoption and dissemination of planting material and improved varieties through breeding mechanism, landrace collections and natural reservoir to be safeguarded in germplasm bank, represent an opportunity to sort out decreasing crop yields and another kind of risks related to pests and diseases. Thirdly, water management spotlights the disjunction between higher-yieldings and water resource consumption. Cassava can grow in areas where rainfalls reach about 400 millimeters (mm) on average a year, however, yieldings respond positively to additional quantities of water supplied by irrigation and rainy seasons. Finally, the development of new improved varieties can respond better in terms of using and applying fertilization to obtain higher-yieldings, even though this crop can resist several dry seasons with poor soils. In addition to being susceptible to diseases and pests as anyone crop, improved cassava varieties do not always perform well with use agricultural pesticides and fertilizers, regardless of the high cost of this type of inputs. Another key characteristic to highlight on traditional and improved varieties is the harvest time. All of our varieties take around 13-15 months to produce harvest. This is relevant for adoption process since farmers can substitute among the seeds easily avoiding high adjustment costs.

On the basis of the prior discussion, a big part of the real concern lies in enhancing and boosting the performance of cassava yield as the Figure 2.3 just displayed. In this sense, the development of improved varieties of seeds is a key mechanism to overcome the barriers associated to low productivity and a set of natural threats that jeopardize the normal functioning of production. In such a way, what institutions can do is to spread the use of improved varieties among the small farmers to make them more likely to adopt new technology. But, what determines the adoption rate of improved varieties? The next section presents a simple model of social interactions that conducts to explain how the individual choice of adopting is related to spillover effects of the social choice.

2.2 The Model

We model the individual decision to adopt improved varieties of cassava seeds in the presence of social interactions by following the baseline model developed in Brock and Durlauf (2001) and Marini (2011). The definition of social interactions refers basically to the perceived utility by an agent given the choices of others in the agent's reference group. This benchmark model is suitable for testing the impact of collective beliefs on individual decision-making since based on a binary choice setting, it can be fit in an empirical framework without incurring in any type of econometric intractability.

The structural model relies on two important dimensions of the economic analysis: the utility maximization and strategic interaction. It can be seen as an optimization problem where equilibrium outcome is the result of a game-theoretic model. In this setting, the social networks which individuals belong to shape the agent's decision seen as a market equilibrium. Depending on the parametric assumptions what we want to impose on, the model could lead to social traps that characterizes multiple equilibria by using other forms of social choices. We opt for the simplest case that exemplifies non-cooperative choices⁴, that is to say, individuals do not coordinate or communicate their decisions. The latter assumption allows us to close the model because the simplification of non-cooperative choice is closely related to the typical equilibria analyzed in games theory⁵.

Another striking characteristic of this model is the presence of complementarities in the individuals' choices. This aspect is related to Banerjee (1992) model as an individual who belongs to a specific group whose majority of people are, for our application case, adopters, is more likely to become adopter than non-adopter. Otherwise, if majority of people are non-adopters, the individual is prone to become non-adopter due to herding behavior (Marini, 2011).

We start by defining a total number of households, I , each of them indexed by $i = 1, 2, \dots, I$. Each household has to choose a binary action denoted by ω_i with support $\{-1, 1\}$. In our case it could be seen as $\{A, NA\}$ where A is the Adoption level choice and NA is Non-Adoption

⁴As mentioned, the assumption of non-cooperative choices is used to estimate a unique solution of the structural model, avoiding any sort of econometric intractability. In addition to this, we are also unable to go through additional kind of behaviors such as cooperative choices since our dataset do not provide information at that level of detail. Testing cooperative choices would imply to iterate over the endogenous variable to recover its structural parameter, which would end up changing the model laid out and considering new source of data. We believe that non-cooperative choice may fit quite well as any type of non-competitive strategy is not likely to see among those farmers who do not participate in a well-developed market.

⁵In terms of social interactions model jargon is the so-called *self-consistency* assumption or a Cournot equilibrium case with N firms in games.

and $A > NA > 0$. We let $\omega = (\omega_1, \dots, \omega_I)$ as the space of all possible set of actions by the population I and $\omega_{-i} = (\omega_1, \dots, \omega_{i-1}, \omega_{i+1}, \dots, \omega_I)$ the choice of all agents but agent i . We assume household utility $V(\omega_i)$ conformed with three components which takes the form:

$$V(\omega_i) = u(\omega_i) + S(\omega_i, \bar{m}_i^e) + \varepsilon(\omega_i) \quad (2.1)$$

where ω_i has been defined above, $u(\omega_i)$ is the private utility associated with the choice, $S(\omega_i, \bar{m}_i^e)$ is the social utility associated with the choice where $\bar{m}_i^e$ is the conditional probability assigned by agent i to the choice of others agents in his group when making a decision and $\varepsilon(\omega_i)$ is the stochastic term or i.i.d term. Therefore, the total agent utility is the sum of a private and social component and a random term.

The social utility term, $S(\omega_i, \bar{m}_i^e)$, follows the expression:

$$\bar{m}_i^e = (I - 1)^{-1} \sum_{j \neq i}^I m_{i,j}^e \quad (2.2)$$

$m_{i,j}^e$ denotes the subjective expected value from the household point of view i of household j choice. In other words, the belief that the household forms about the group choice is the expectation of the remaining beliefs of the other households that belong to the group rather than considering his own expectation.

We impose parametric assumptions on the functional form of the social utility term. The election of the following specification is based primarily on Cooper and John (1988) in order to capture strategic complementarity effects. This means that we are interested in effects such as:

$$S(\omega_i, \bar{m}_i^e) = J \omega_i m_i^e \quad (2.3)$$

$$\frac{\partial^2 S(\omega_i, \bar{m}_i^e)}{\partial \omega_i \partial \bar{m}_i^e} = J > 0. \quad (2.4)$$

The parameter J measures the degree of interdependence across agents. The functional form in (2.3)⁶ allows to assess through J how the decision taken by the household and the collective belief interact each other, in addition to providing much more easily for the purpose of econometric tractability. J is also known the "social multiplier", which

⁶Notice that the functional form can vary to assess non-linear effects in order to analyze multiple equilibria under social traps conditions. As a matter of fact, the equation (2.4) can be derived from the functional form seen in (2.3) and the quadratic form $S(\omega_i, \bar{m}_i^e) = -\frac{J}{2}(\omega_i - \bar{m}_i^e)^2$. The quadratic form penalizes deviations from the mean more strongly than the proportional spillover case (equation 2.3). A simple manipulation of the quadratic form shows that this one and equation (2.3) are equivalent, differing on whether are levels or deviations. See Brock and Durlauf (2001)

points out how household decision, is shaped by the specific group which belongs to. J reflects a multiplicative interaction between the individual decision and the social utility term. According to Marini (2011) and Brock and Durlauf (2001), as long as a sizeable social multiplier is obtained, the presence of multiple equilibria is more likely to have. J is our target of interest on empirical ground.

We partially complement the baseline model already laid out with Zanella (2007) and Marini (2011) approach as the bridge to focus on the empirical implementation. Let us suppose that the equation (2.1) takes the form:

$$V_i = v_g + h_i\omega + J_g\omega m_{ig}^e + \varepsilon_{ig}\omega + \varepsilon_{ig} \quad (2.5)$$

where g represents that the household is a member of one of the group⁷ $g = 1, 2, \dots, G$; and v_g is the net profit ($v_g = \kappa_g - \rho_g$) where κ_g and ρ_g ⁸ are benefits and costs of belonging to a group, respectively; $h_i\omega$ is the private deterministic utility, $J_g\omega m_{ig}^e$ represents the peer or neighborhood effects. The terms $\varepsilon_{ig}\omega + \varepsilon_{ig}$ is the stochastic component capturing the random private utility where each of its components represent the unobservable disposal to make a choice given a membership of a group and the unobservable disposal to choose a group g to live.

We assume households facing sequential choices which indicate whether they decide first the group to live with after their behavior and vice versa or decide simultaneously the group and the behavior to choose⁹. The optimization problem requires to be solved backward in two steps: in the second stage, households choose their behavior given their membership into the group and the distribution over resting households choices. When assuming self-consistency, the probability distribution corresponds to form expectations on average behaviors. The optimization problem faced by the household is:

$$\omega_i = \sup_{\omega \in \{-1, 1\}} \{h_i\omega + J_g\omega m_{ig}^e + \varepsilon_{ig}\omega\} \quad (2.6)$$

After optimizing the maximization problem, in the first stage, households solve the problem associated to choose a group and taking into account their behavioral choice given in the second stage. The optimization problem is given by:

$$g_i = \sup_g (v_g + \varepsilon_{ig} + \sup_{\omega} \{h_i\omega + J_g\omega m_{ig}^e\}) \quad (2.7)$$

⁷In our sample, g symbolizes the set of municipalities in the State of Cauca, Colombia.

⁸ v_g can be understood as fixed effects at group level.

⁹According to Zanella (2007) and Marini (2011), it is not relevant the way of solving this game since the outcome leads to a symmetric solution.

Like in all theoretical analysis of social interactions, we assume $\varepsilon(-1) - \varepsilon(1)$ are independent and extreme-value distributed or sigmoid distribution function. The parameterization of $\varepsilon(-1) - \varepsilon(1)$ allows to calculate $Prob(\omega)$ since $V(-1) > V(1)$. The difference in $V(-1) - V(1)$ yields the density function for the likelihood of adoption. The logistic functional form derived from this specification conducts to modeling the individual adoption choice as a function of a set of variables. In other words, the outcome of the parametric distribution leads to the following result:

$$p_{i\omega|g} = \frac{\exp(h_{i,g} + Jm_g^e)}{\sum_v \exp(h_{i,v,g} + Jm_g^e)}$$

$$\omega_i = f(k, X_i, Y_{ig}, m_{ig}^e, \varepsilon_i, \varepsilon_{ig}) \quad (2.8)$$

The household choice, ω_i , is function of observable household characteristics, X_i ; observable exogenous characteristics at group level, Y_{ig} ; the endogenous effect or the household behavior influenced by the group behavior, m_{ig}^e and ε_i is the stochastic term attributed to unobservable household characteristics. k , X_i and Y_{ig} put together the private utility, h_i and m_{ig}^e the component of social utility. The next section presents the econometric framework and the discussion on the key challenges found in empirical literature.

2.3 Econometric Approach

A main strength of the simple model developed here is its correspondence with the classical discrete choice literature or Random Utility Model. The advantage of this benchmark model is the structural meaning of its parameters which, in principle, can be computed to analyze policy changes, as well as being capable of taking into account striking features such as multiple equilibria and phase transitions not considered in models of the type linear-in-means (Brock and Durlauf, 2001; Brock and Durlauf, 2003, Brock and Durlauf, 2007). Based on the result achieved in (2.8), we estimate a binary discrete choice model. We assume once households choose the group $g = 1, 2, \dots, G$, they make a choice regarding their behavior $\omega \in \{A, NA\}$. Thus, we are interested in estimating the specification:

$$p_{i\omega|g} = \frac{\exp(c'X_i + d'Y_g + Jm_g^e)}{\sum_v \exp(c'X_i + d'Y_g + Jm_g^e)} \quad (2.9)$$

Empirically, ω represents the binary choice whether a household adopts new technology or not, $h_{i,g}$ is the private utility captured through specific characteristics at household level, X_i , and covariates at municipal level, Y_g . The social interaction term which we are interested

in, m_g^e , symbolizes the expectation of household i about the average choice of the other households belonging to the chosen municipality. The use of the self-consistency assumption, that is, $m_g^e = \bar{m}_g$, allows to close the model indicating that each household can form a rational expectation about resting households' beliefs, by assuming every household has complete information of the individual characteristics of the others households in each municipality. In other words, we match the objective probability with the average adoption of the municipality. In this way, households decide to adopt or not based on average level of adoption they expect to see in the municipality.

The specification in equation (2.9) has been the workhorse in empirical work. In fact, the detail level of data required for testing prediction derived from this kind of models has prevented the flourishing of this analysis on empirical ground. Brock and Durlauf (2001) point out that implementing an empirical analysis requires individual level observations rather than using aggregate datasets (Conley and Topa, 2000), like in most other studies carried out. On this, there are additional concerns over two identification problems widely recognized in this literature: correlated non-observables and the reflection problem (Manski, 1993; Brock and Durlauf, 2001; Bandiera and Rasul, 2000). As Bandiera and Rasul (2000) have stressed, the typical regression estimated is a linear version of (2.9):

$$\alpha_{iv} = \beta X_i + \gamma \alpha_{n(i)v} + \delta X_{n(i)v} + u_{iv} \quad (2.10)$$

In most part of literature, correlated non-observables refers to the omission of covariates that are specific to the environment or geographical position, that is, a much simpler version of (2.10):

$$\alpha_{iv} = \beta X_i + \gamma \bar{\alpha}_v + u_{iv} \quad (2.11)$$

Generally, individuals do not report information on belonging to specific groups or communities and observing such specific characteristics at that level may be tricky, unless the objective of analysis has previously been identified to apply this approximation to a particular problem or data.

The most common empirical analysis relates the group or neighborhood to the geographical location due to being relatively easy to observe. However, the specification that primarily ending up estimating is (2.11) because not in all the cases geographical characteristics allow to enrich the part of the dataset attributed exclusively to individual characteristics. Hence, the correlated non-observable arises at the moment of estimating (2.11). The consequence of using (2.11) without controlling for specific covariates at geographical level is the bias in the social interaction term since, as the social interaction is reduced to municipality means, the correlation between this one and municipalities characteristics becomes stronger. This

implies that the positive effect of γ (or J in our specification) can not be interpreted as a measure of the social impact on individual propensity to adopt and its movement is due to omitted or unobservables covariates. Our analysis does not suffer completely from this problem because we count on information to mitigate the bias in the term of interest.

Alluding to the reflection problem treated in Manski (1993), this is a bit more challenging to overcome in practice. The reflection or simultaneity problem refers to the idea of symmetry in the social interaction term, namely, household's behavior affects the behavior of the other households of the group and vice versa. Measuring the impact of this identification problem requires data at individual social network level rather than aggregate measures of networks¹⁰. We try to spotlight empirically the importance of this issue, by measuring closeness and betweenness as of geographic distance social network that we construct.

¹⁰Bandiera and Rasul (2000) derives mathematically the simultaneity bias by using a complete or symmetric network. The size of the bias tends to disappear depending on the size of the network.

3. Empirical Analysis and Results

3.1 Data Set and Sample Description

Our analysis of adoption uses different datasets that come from two sources. We focus on studying the adoption rate of modern improved cassava varieties in the State of Cauca in Colombia. The study area is located in the South West of the country, covering a wide variety of thermal floors and mountainous area, being the most propitious for crop diversity. The State is conformed by 42 municipalities whose Capital City is Popayán. Cauca has been home to rebel groups that have promoted armed internal conflict for many years, boosting illegal crops production and affecting market opening due to risk of insecurity in the area. In this sense, the aim of encouraging farmers to adopt new seeds is not only devoted to improving households productivity conditions but also for replacing illegal crops in order to contribute to reducing of high levels of crime and violence.

We use the first data set that comes from a pilot project headed by the *International Center for Tropical Agriculture* (CIAT) in collaboration with *Universidad Nacional de Colombia*. The process of filling out surveys to 293 farmers was administered during three months (November and December in 2014 and January in 2015). A subsequent and extensive data cleaning process was implemented to correct certain errors that were detected at the moment of filling out the question paper, among others. The sample consists of socio-economic characteristics which includes several kind of information at different levels¹ We covered 18 municipalities² in the survey since some of the additional locations that we wanted to sample were not able to contain because of security risks in the area.

Our first set of covariates includes variables at the household level as described in Table A.1 Some variables are provided directly by the survey while others were constructed as in the case of *Appliances*, *Household size*, *Head of household' scholarship* and *Head of household's*

¹The survey includes information at the household, plot, variety level, among others. However, we decided to reduce all of the resting information levels to household level to reach better measurements as well as applying to the base model.

²The municipalities sample are: Buenos Aires, Cajibío, Caldone, Caloto, El Bordo, El Tambo, La Sierra, Morales, Padilla, Piendamó, Popayán, Puerto Tejada, Rosas, Santander de Quilichao, Timbío, Villa Rica.

age. We also constructed the outcome variable which labeled *Adoption* in the database by following the standard measure in literature, that is, as an indicator or binary variable which takes the number 1 if a household adopted an improved seed and 0 otherwise. This approach has specially been criticized by Feder et al. (1985) for being a common misconception about certain adoption rate. The authors point out that *intensity* of adoption is much more important than computing a simple "adoption" or "non-adoption" label. This is due to the fact that it is not the same considering a farmer who is using 1% or 90% of HYV's in his hectareage. We slightly avoid this problem for two reasons: our dataset reports area in hectares at the farm and crop but not at the variety level. This obeys the fact that farmers have a tendency to mix several kinds of crops and their varieties of them for reasons that have to do with agricultural and agronomic purposes³. Furthermore, the improved cassava varieties trade is not still a wide developed market in which farmers may respond to positive demand shocks in the sense that they allocate land area or specific plots to varieties in high demand.

The genetic identification process of improved cassava varieties had two levels of analysis. Firstly, during the data collection the team in charge of filling out the surveys asked the farmers to report the number and name of varieties that had in the farm. Thus, the team wrote down this information together with the rest of socio-economic data. Secondly, the team of geneticists took samples of the varieties that had passed the crop cycle. In many cases, the sample material was not available since it was under sowing period. Altogether, the geneticist managed to survey 210 households with 450 sample available to test in the lab. The total number of genetic groups identified on the basis of the samples were about 30, some of which were labeled as improved varieties. Figure 3.1 depicts the dendrogram for the set of samples collected and its genetic group. Initially, many mistakes were made around the classification of some groups which were classified as improved varieties when they actually were not. For instance, the group G-01 which belongs to the starched-group and named for some farmers as "algodona", "algodón", "algodonosa", "algodonera", was misclassified as improved variety. A series of successive test showed the opposite result. On the Figure 3.1 the CIAT slogan indicates the genetic group that is an improved variety. In order to seize the larger sample size provided on the farmer's socio-economic data side, we compute the correlation between the geneticists' sub-sample that exhibits the set of improved varieties analyzed genetically and the identification made by the farmers, leading to a value of 0.40. In the latter case, we labeled the names reported by the farmers as improved variety by keeping track of the unusual names that were highly likely to match with a numeric code. For example, names reported such as "p11", "f40", "ica", "cm7138-17", "-7", among others,

³When mixing different types of crops in an specific plot or farm, farmers can take advantage of mineral nutrients from land and crops to improve the yield as a whole.

corresponded to improved seeds.

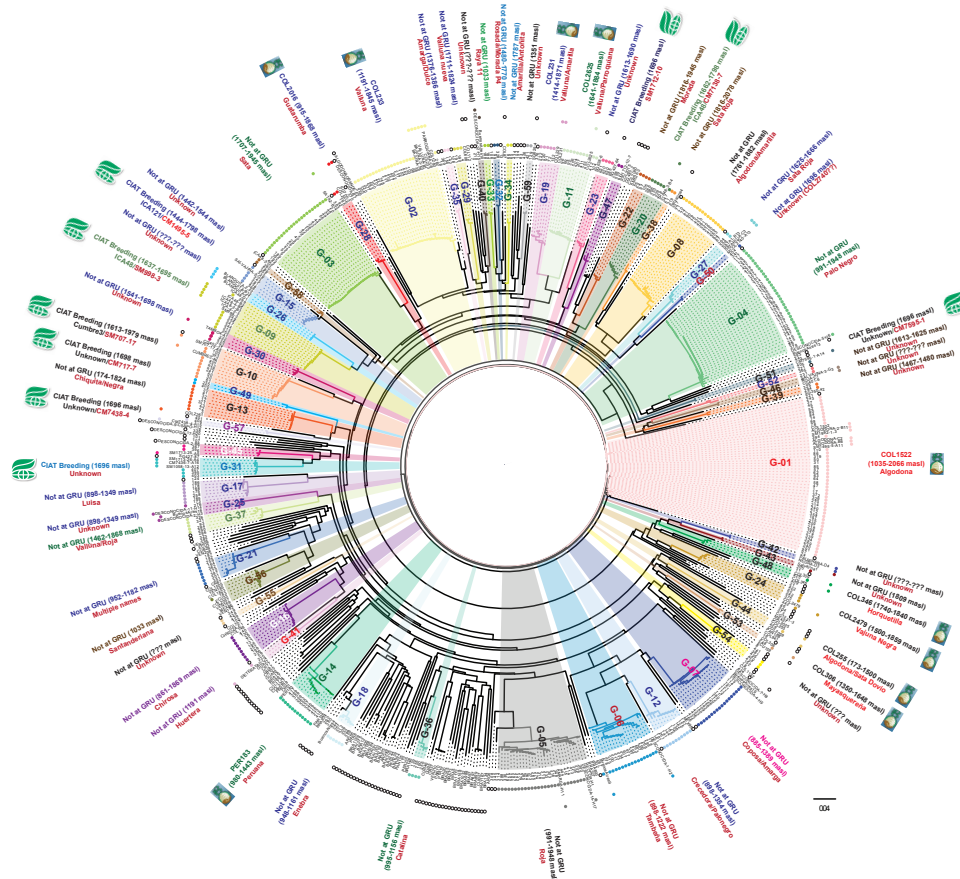


Fig. 3.1 Source: CIAT.

Our second set of covariates is taken from the *Sistema de Información Geográfica para la Planeación y el Ordenamiento Territorial* (SIG-OT), a web page⁴ linked to governmental entity *Instituto Geográfico Agustín Codazzi* (IGAC). This source of information provides a number of socio-economic data for municipal and departmental geographical units in Colombia. SIG-OT works with *.shp* extension file or *shape* file that hosts geographical coordinates or polygons, accompanied by a set of *.prg*, *.dbf*, *.sbn*, *.sbx*, *.shx* extension files that allow to project the geographical patterns on the space. We merge our database with those extension files by removing character errors found in primary keys.

We choose the main variables suggested in literature like human capital index and poverty conditions and added additional regressors of interest available in the database. Unfortunately, our sample did not match with variables associated with violence and armed conflict

⁴Web page http://sigotn.igac.gov.co/sigotn/frames_pagina.aspx

such as crime rate and anti-personnel mines number because of missing values and lack of variability⁵.

The Human Capital Index that we take is a computed score provided by the *Departamento Nacional de Planeación* (DNP) thorough SIG-OT in 2005. The Unsatisfied Basic Needs is a percent measure provided by the *Departamento Administrativo Nacional de Estadística* (DANE) in 2010. We also added a Quality Life Index rural score measure, the Housing Quality rural score and the Phone Coverage percentage available for 2005. Road infrastructure and land conflict variables were reported without information in the files.

Satellite data of geographical positions were retrieved from **Google Maps api key**⁶. We connected to the website from the R terminal. The web-page allows to measure the distance between destination and origin points, as well as the transportation time, the road status and transportation mode. The api sets a restriction over the number of points to enter, offering a free access up to 100 distances stored in a distance-matrix. Additional queries require payment for unlimited access. Therefore, we divided the data into subgroups by means of clustering. We carried out a web-scraping⁷ process since the information is presented in XML or JSON format. We extracted the distance and time in terms of the driving mode, by assigning the origin point as a non-adopter and the destination point as an adopter.

⁵The antipersonnel mines number that only matched with our municipality sample was zero

⁶For more information <https://developers.google.com/maps/?hl=es-419>

⁷Web-scraping is a computational technique which consists of extracting information from websites.

3.2 Empirical Measures of Social Interactions

As it has been mentioned previously, the most common measure to approximate social effects is the simple village average in the group which a microeconomic agent belongs to. The problem of this approach is that we are not sure whether the estimated coefficient really captures a social effect when omitting covariates at the group level. In any case, we compute the average social effects in the regression results—in other words the equation 2.2—by adding the specific controls already described.

In order to provide more evidence on existence of social interactions, we construct a new variable which captures the degree of interaction between adopter and non-adopter households through geographical distance. The geographical location is another way of approaching to relationship between neighbors and therefore learning process (Jackson, 2008; Brock and Durlauf, 2001). For instance, it can be considered that agents interact each other not only through physical distance but also through political belief, gender and race closeness, among other things. However, for our purposes, the geographical location is a reasonable measure to assess the impact of social effects on probability of adoption. This measure is used because we have a farmers sample who live, on average, under monetary poverty conditions, in which the only way to communicate with their friends and family is through physical contact. In this case, we do not expect farmers to discuss about adoption process by means of a cellphone. Figure A.1 shows the spatial distribution of adopter and non-adopter households located across the State where the red circle indicates a household that did adopt an improved seed of cassava. According to Table A.1, the percentage of adopter households was 18%. We want to investigate how the probability of adoption is affected when a non-adopter household strongly interacts with an adopter household (the degree of closeness in terms of distance is much stronger). Toward that objective, we apply a machine learning clustering algorithm in order to make groups of households for computing the minimum distance from a non-adopter to an adopter household. This algorithm works like calculating a search radius of a geometric figure. We opt for applying this algorithm instead of using a specific radius function that provides us more flexibility because some municipalities had adoption rate of zero percent as shown in Table A.1. Thus, increasing the radius of action would be interchangeable with expansion provided by the algorithm.

One particular advantage of the database is that our households are geo-referenced so we exploit this information that is not commonly observed in this type of analysis. As a matter of fact, the use of a geo-referencing system is much more suitable for achieving robust measures of equation 2.2 due to we are able to form a social distance matrix as equation 2.2 suggests in the term $m_{i,j}^e$. We implement the standard k -means clustering algorithm widely

used in statistical literature on machine learning. Briefly, we have a set of observations $(x_1, x_2, \dots, x_n)$, with n as the sample size. k -means partitions the full sample into k sets $\mathbf{S} = (s_1, s_2, \dots, s_k)$ with $k \leq n$, so the objective is to minimize the distance of each point to the K center. Mathematically, the objective function is:

$$\arg_S \min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (3.1)$$

where μ_i represents the mean of observations in S_i . The algorithm uses an iterative method in two steps. It sets a set of k -means $m_1^{(1)}, \dots, m_k^{(1)}$. In the first step or assignment step, the algorithm assigns each observation to the cluster in which the sum of square is minimized or the observation is the nearest mean. Formally,

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \leq \|x_p - m_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\} \quad (3.2)$$

where x_p represents an only assignment to $S^{(t)}$. In the second step or update step, the algorithm computes new means in order to be the centroids of the observations in the new recalculated clusters:

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad (3.3)$$

Figure A.2 plots the number of clusters that we obtained as result of applying the k -means algorithm. The number of clusters obtained are appropriate because we count, at least, with one adopter in each group. Once the algorithm has firmly been in place, we proceed to construct a distance social network for each cluster by using the Euclidean distance matrix specified as:

$$\delta^{euc}(x_1, x_2) = \sqrt{\sum_{i=1}^n (a_i(x_1) - a_i(x_2))^2} \quad (3.4)$$

The equation 3.4 yields the distance between two points without scale or measurement unit because we used the latitude-longitude points. During the data cleaning process, we had to re-scale the geographic coordinates since they were originally in WSO84 system. In order to facilitate the interpretation, we programmed a functional algorithm that undertakes several tasks at the same time. The algorithm computes the distance in kilometers by road through connecting to Google Maps API key. Then, the algorithm implements a searching of each non-adopter in each cluster and computes the minimum distance in kilometers to the adopter household. If the algorithm detects an adopter household, assigns simply zero. Figure A.3 displays the distance distribution of non-adopter with respect to adopter. From the regression

analysis in the next section, we expect the variable to be statistically significant and inversely correlated with the probability of adoption.

3.2.1 Closeness and Betweenness Measures

We provide additional measures of social effects based on some Social Network techniques. In the jargon of this line of analysis, we basically count with an adjacency matrix which is characterized by being a complete weighted undirected network structure. According to Jackson (2008), we formally define a *graph* (N, g) as a set of nodes $N = \{1, \dots, n\}$ and a real-valued $n \times n$ *adjacency matrix*, g . This matrix represents a list of nodes linked to each other. A *weighted* graph or network refers to the degree of intensity of level of relationships among the entries of g . On the contrary, if the value of the entries are 0 or 1, the network is *unweighted*. On the other hand, it is said that a network is *directed* if $g_{i,j} \neq g_{j,i}$, otherwise, $g_{i,j} = g_{j,i}$ is an *undirected* network.

Networks can be described through the use of "macro" measures that set global patterns, looking at the broad characteristics of the graph and "micro" measures that emphasize on the relevance of nodes positions, respectively. We conduct the analysis on the last ones by computing two standard measures of centrality. The Closeness Centrality determines how easily a node can reach other nodes. It is just the inverse of the average distance between i and any other node. Formally,

$$CC = \frac{(n-1)}{\sum_{j \neq i} l(i, j)} \quad (3.5)$$

where $l(i, j)$ is the number of links in the shortest path between i and j . The second measure of centrality is the Betweenness Centrality which represents a node as an intermediary or connector. Let us define $P_i(kj)$ being the number of geodesics or shortest paths between k and j that i lies on and $P(kj)$ the total number of geodesics between k and j . The insight of this lies in analyzing how important is i in terms of connecting k and j by looking at the ratio $\frac{P_i(kj)}{P(kj)}$. The closer the ratio to 1 is, the more important i becomes to link k to j . Otherwise, when the ratio is close to 0, i is less critical to k and j . We get the betweenness centrality by averaging across all pairs of nodes, that is to say:

$$Ce_i^B(g) = \sum_{\substack{k \neq j \\ i \notin \{k, j\}}} \frac{\frac{p_i(kj)}{P(kj)}}{\frac{(n-1)(n-2)}{2}} \quad (3.6)$$

Unlike the minimum distance in kilometers that we exposed, the current centrality measures do not discriminate or compute with respect to a specific adopter or non-adopter as the

previous variable does. So, we may expect closeness and betweenness to explain how the role of interaction and the position are relevant in terms of affecting the propensity to adopt.

3.3 Regression Results

The preliminary analysis of the data shows a remarkable pattern of the adopter and non-adopter households located across the space (Fig.A.1-A.2). The points are distributed in a way that is suitable for measuring the information channel between adopter and non-adopter through the road distance, since the number of adopters is small. This could be interpreted as the influence an adopter can have on the non-adopters to force the adoption process. The implementation of the k -means algorithm on the spatial distribution in Fig.A.1 leads to a randomized group conformed by three clusters (Fig.A.2). The analysis of the clustering allows us to have at least one adopter by group. As it was explained in the previous section, we compute the minimum distance in kilometers by road from a non-adopter to an adopter for each cluster. Fig.A.3 displays the histogram and density of the variable generated as a result of the algorithm. In general, a high percentage of non-adopters are between 0.1 to 10 kilometers away from the nearest adopter. The shape of the distribution resembles the negative exponential law, so we expect an inverse effect from the regression analysis.

On the other hand, we also implement the standard analysis of social interactions by looking at village adoption rate. Table A.3 shows the adoption rate by village. Of the total village in our sample, 9 municipalities had an adoption rate greater than zero, being Padilla and Morales the municipalities with the two higher rates of adoption.

We start by explaining the first regression analysis of Table A.4 presenting the effect of village adoption rate on the probability of adoption. Column (1) shows the effect of the average village adoption rate on the individual decision to adopt without including individual and village characteristics. As the theory suggests, the sign of the effect is expected to be positive. In other words, as the average adoption rate increases, the individual household decision to adopt increases. Column (1) exhibits the expected result and statistically significant. Column (2) adds only controls at the household level. In this case, the size of the estimated coefficient reduces (although the difference between coefficients is not significant) and also continues statistically significant. Column (3) incorporates additional controls at the municipal level. Before analyzing the results as a whole, we add the additional column (4) that controls by village fixed effects. The reasons for including this column and excluding the variable of interest lie in comparing the differences in the size of the estimated coefficients and avoiding econometric problems related to multicollinearity. This problem arises because the fixed effects and the average village adoption rate capture the same, so the degree of correlation is

potentially high. In other words, village fixed effects and geographical variables constructed from geographical patterns exhibit high degree of correlation. Therefore, we refrain from including both variables at the same time and in the following regression analysis. The results of column (3) establish that the adoption process is mainly driven by the strength of social effects and technological variables. The social interaction term works as a mechanism of spreading or *spillover* that forces the household individual decision to adopt. According to the definition of social interactions given in our base model in which agents tend to make decisions based on the average collective behavior that they observe (like following a herding behavior), the estimation results suggest that: farmers who live in municipalities where adoption rate is higher are more likely to adopt. This estimated coefficient can also be interpreted as a means of spreading of public information within the village that farmers take into account. Unfortunately, we do not have information of the agents' social networks related to which or how many of their friends, family members or neighbors have influenced their adoption process through simple information exchange. However, we try to spotlight the role of private exchange information. by computing the distance between a non-adopter and an adopter in the next table. Regarding the technological variable, *Technical access* refers to talks on agricultural extension services in cassava received in the last year. This result can be read as the farmers acquire new skills and knowledge about better agricultural practices, the propensity for adoption also increases. These two variables have a meaningful position in the economic analysis since they exemplify the role of interactions and knowledge on productivity. We also find that the variable *Head of household's scholarity* remains statistically significant under each of the specification. This results is quite intuitive since as the level of education for a member of the household is higher, the propensity for a new technology also increases. This explains a close relation between productivity and education, highlighting the importance of having more educated farmers to force new farmers into adopting new technology.

Table A.5 presents the results for our additional measure of social interaction. As before, the three first columns analyzes the effects with and without adding controls. The variable *Distance in km from the nearest adopter*⁸ from column (1) does not have the expected sign and is statistically non-significant. The next two columns specify the estimated coefficient under the respective covariates at household and municipal level. The full outcome in (3) yields the expected sign with a statistically significant result. The insight behind this analysis suggests that as the minimum distance in kilometers from a non-adopter to an adopter increases, the probability of adoption decreases. In other words, if a non-adopter moves away

⁸The results presented for this variable take the average at the village level. The key difference with regard to Table A.7 is the variable has been computed at the housing level.

from his nearest adopter neighbor (the minimum distance between them goes increasing), his probability of adoption goes reducing. This is the most striking evidence on approaching to spillovers and learning by neighborhoods effects since they are also framed in the context of learning and diffusion process in industry.⁹ Secondary outcomes from the regression allow to also see how the effect of *Technical access* remains stable. Moreover, under this characterization, the controls at the village level become meaningful. For instance, the human capital index affects positively the probability of adoption, while the Unsatisfied Basic Needs Index (UBN) reduces the probability. This result may be indicative with that of the state of poverty which hinders the access to new technology. In fact, it is not a strange result if we consider that poverty set barriers for farmers in developing countries.

Table A.6 shows the same results found in Table A.5 varying the transportation mode in lieu of the Euclidean distance. We ruled out the regression with the two variables in order to avoid collinearity. According to the table, the sign of the effect and its statistical significance remain stable, although the size of the coefficient lowered. This is consistent with the idea that the Euclidean distance overestimates the effect as to driving mode because the Euclidean measure is a shorter distance in a straight line. This comparative is useful since provides an insightful view of how shorter distances impact positively the propensity to adopt.

Our previous results provide evidence of social effects as the theory has shown. Nevertheless, it remains unclear whether the adoption is driven by a learning process. For instance, a non-adopter farmer can adopt new technology by the fact that his closer circle of friends or neighbors do the same. But at the same time, he may refrain from using new technology if he does not know how to implement and apply the management of the improved seeds in technological terms. A farmer is more likely to adopt if his neighbor has knowledge of agricultural practices. We intent to identify this effect by conditioning in a farmer who is trained in extensive services and other use of technology. Column 2 of Table A.7 estimates the distance from the non-adopter to the nearest trained adopter. The results for this case suggest that learning by neighborhood may be more related to closeness social effects than technological effects. In other words, farmers are willing to adopt if someone close to their neighborhood does. The size of the estimated indicates the importance of the neighbors regardless of a neighbor is trained.¹⁰ However, when conditioning for use of agronomic techniques such as weed control and fertilizers, the estimated coefficients are similar to the social effect based on physical distance. Columns 3 and 4 show the minimum distance from the non-adopter to an adopter who used either a weed control technique or fertilizers,

⁹Carvalho and Voigtländer (2014) have a similar version of our approach applied to input-output linkages but using network proximities thorough the minimum-distance algorithm known as Dijkstra's shortest path algorithm.

¹⁰Another effects are susceptible to be present. We discuss additional concerns in the section 3.4

respectively. Unlike the Column 2 that makes emphasis on informative talks on agronomic management and experiences associated to cassava crop, Columns 3 and 4 stress the effective use of agronomic techniques into the hands of farmers. In other words, it is likely that the estimated effect might be learning rather than a mere social effect. On this, it is critical to mentioning that one side of the effect could be a social effect in the sense mentioned before as a farmer who is adopting because his neighbor doing, or because the farmer is effectively learning. Part of the future analysis requires the construction of new mechanisms that allow to disentangle this identification problem on data. Finally, we proxied an imitation effect by computing a dummy which captures a neighborhood of non-adopter households close to an adopter household with outlier-characteristics. Among the characteristics, we identified adopter with high level of education, aged and at a distance between 0,02 and 2,16 kilometers. As in Bandiera and Rasul (2001), the impact resulted positive but statistically non-significant. On this point, we are also aware of the necessity for more appropriate measures that take into account imitation effect from data.

The next two tables asses and discuss the impact of two descriptive networks measures that we constructed as of the households distance social network: closeness and betweenness degree. Both of them were set as complete weighted undirected network structure where the edges between households represent the distance in kilometers. A clear disadvantage of this approach is the measures do not focus on one side of the agents as compared to the previous measure analyzed on Table A.5. So those effects must be interpreted as global effects and not related to a specific group of households.

The first set of results is found in Table A.8. Column (1) shows an inverse and statistical significant effect of closeness degree on probability of adoption. Nevertheless, the coefficient estimated sign remains unstable since we would expect a positive effect, that is to say, as the closeness degree increases, the adoption rate also increases. In other words, the higher degree of interaction is, the more probability to adopt is. In column (2), the sign and statistical significance of the coefficient behave well when adding controls at household levels. The positive effect may be interpreted as a mode of transmission from the adopter to a non-adopter to adopt. Finally, when adding controls at village level, the coefficient fails in terms of the expect sign and statistical significance. This evidence sheds light on the relevance of controlling characteristics of the group in which agents belong to, as it was highlighted by the mentioned before.

The results for the betweenness degree are found in Table A.9. In this case, we would also expect a positive effect on the adoption rate since, as the position of a node with regard to the other nodes becomes stronger, the resulting effect of the position in the network on the adoption rate also increases. This insight is closely related to the work of Beaman

et al. (2015) in the sense of using some trained targeting households to spread across their neighbors, the adoption process. Overall, the estimated coefficient has the expected result without including controls. However, the coefficient becomes statistically non-significance under household and village controls with the correct sign. This situation allows to analyze the case when the variable does not pass any of the controls. Finally, Table A.10 presents the empirical results for the model laid out in section 2.2, showing the logistic regression for our variables of interest *average village adoption rate* and *distance in kilometers from the nearest adopter*. Column 1 shows the estimated coefficient for the parameter J or social multiplier. The estimation suggests a sizeable impact of social interactions on individual adoption choices. The meaningful size of the estimated parameter could be explained as the overall and spillover effect that the adopter-regions (membership to regions) have on both adopter and non-adopter farmers. In other words, the propensity to adopt a new technology counters the effect of non-adopters' non-adoption choices. This sizeable impact could also be understood as the close proximity a set of adopter and non-adopter farmers share each other into the village. Due to household farmers interact across short distances, the size of the coefficient is much bigger. On this, for her study on cultural beliefs, Marini (2011) reports a J of about 1.40 and 1.60, notwithstanding, the geographical positions that she used are regions from the United States, which are much further away among them in comparison with our villages of analysis. Bandiera and Rasul (2001) present a social multiplier but in a linear-regression model. In this way, we are not able to set a similar analysis between the models. However, the coefficient that they report is 0.482, much lower than our estimate of the linear-version.

So far, the analysis only represents a part of the social interaction effects. For instance, we could also be interested in getting into the aggregate and individual network measures as Bandiera and Rasul (2002) and Beaman et al. (2015) examine. Our analysis is a part of the story since agents can develop more complex adoption process on the basis of the physical distance analyzed here. The closeness in terms of family or community ties plays a meaningful role to strengthen the links between adopter and non-adopter in order to boost the adoption of new technology. This task also requires better measures of variables at group level that allow to control for correlated non-observables and simultaneity which could lead to unreliable results.

3.4 Discussion

According to the behavioral model that we have chosen and the empirical evidence provided, the adoption process is significantly determined by the size of social interactions. By social interactions we understand how the agent's choice is influenced by the average collective decision that the agent observes. Due to the parametric assumption of complementarity, the agent is more likely to make a decision based on the main choice that the rest of his neighbors have taken. This assumption spotlights the role of herding behavior and the degree of closeness widely recognized in theoretical models that agents experiment when relating to each other. For our application case, technology adoption in agricultural economics is mainly driven by the force of learning as theoretical construct. Learning by neighborhood, as one case of learning, is associated to how farmers take advantage of their neighborhood through the exchange information to know more about the new technology. For instance, some talkative farmers convey their ideas of new improved seeds allowing to speed up the process of adoption. Meanwhile, some other farmers merely adopt by following the same behavior of their friends or neighbors. We are unable to verify more appropriate measure in the information channel since we do not observe directly in the dataset. One feasible question to keep in mind in future analysis would be to ask farmers for how talkative they are with the rest of their farmers in order to share their ideas and techniques on agronomic practices.

The empirical evidence provided in the previous section is highly suggestive of the presence of peer effects (social interactions). Given the wide set of covariates by which we have controlled, we can almost certainly say that the estimated effect is a causal relation. Based on the theoretical framework developed, we read the results as follow: On the one hand, the average village adoption rate is indicative of farmers are likely to become adopters since they are expected to make a decision on the majority choice observed in a region. In other words, small-holder farmers base their choice regarding the higher average rate in between the binary choice. If a farmer expects the region which belongs to has a higher rate of adoption than that of non-adoption, he will certainly adopt. There are two important insights behind the latter result: herding behavior and a public information channel. In respect of the first one, farmers set ties with those who share a higher likelihood to behave in a predominant way. They embrace the behavior mostly seen. Regarding the second one, it would also seem that behind herding behavior, it would have a public information channel that allows farmers to know about the spreading of information in terms of the use of new improved seeds. It may be possible that agencies associated to the government stir up information offices through the local government in different parts of the village. We do not have information at detailed level to make sure the real mechanism that differentiates between both forces.

However, we are inclined to believe that social effects such as herding behavior determines the willingness to adopt since the presence of governmental institutions are scarcely present in the zone. On the other hand, we reinforce the idea of social interactions through measuring the distance between adopters and non-adopters. As we have clearly seen in the regression results, the neighborhood effects matter to determine the adoption process. Being around adopters make non-adopters more likely to behave like them. As we said before, we ignore the exact mechanism behind as the information at network level is unknown but we strongly believe that the behavior of the nearest adopter influences on non-adopter to behave in the same way. The lack of a mechanism also precludes us from drawing inferences of learning behavior for certainly. We take for granted that learning by neighborhood implies social interaction but social interaction does not imply learning by neighborhood. On this, the key to differentiating both of them lies in the main mechanism which leads to the set of problems we are going to talk about later on. The fact that two farmers (non-adopter and adopter) interact each other does not necessarily mean learning. For instance, it is possible that even observing the agents' social networks we rule out learning. If the non-adopter's nearest neighbor (either my relatives, friends, members of the church, political party, etc.) is using improved varieties, the non-adopter may refrain from adopting improved seeds as he does not know how to implement technically those kind of seeds. A feasible mechanism of learning may arise when exchanging information between the sides. We suspect the presence of learning in our approach even though we lack of much more detailed information to corroborate the effect. By using the variable of extensive services which refer to talks about management and technical assistance in cassava, we condition the minimum distance in an adopter-farmer who participated in technical assistance. By doing so, we try to approach an information sharing that allows non-adopter to learn about the management of new seeds. In this sense, we would capture the effect of learning given the information that we have. The evidence shown is also suggestive of learning effects. Namely, the estimated effect remains stable and statistically significant.

Additional concerns have to do with *imitation effects* and *endogenous membership*. It is likely that farmers decide to adopt because they just want to imitate the behavior of the best and brightest farmer who lives near them to show off. The most talented farmer acts as a relevant hub in the network of neighbors which is required to be controlled for his specific characteristics. The inclusion of this term would be able to separate the effect that accounts for imitation from that of information sharing and social learning. On the endogenous membership side, the individual networks that we do not observe could lead to much more detailed subgroups of farmers that belong to particular communities such as membership to a club, fund or church. Those sort of memberships also need to be controlled for characteristics

at group level. In this regard, the key to differentiating alternative measures that may have an impact on social interaction term is to carry out questions that go over the farmers' relational environment. Future analysis needs to draw attention to farmers' social networks-specific information.

4. Concluding remarks and future research

This document has studied the relationship between technology adoption decision and social interactions. Based on the canonical model with non-cooperative choices, the empirical implementation supports the main predictions derived from the theoretical analysis. Using a rich set of covariates at the household and village level on modern improved cassava varieties for the State of Cauca in Colombia, the econometric results set that as the average village adoption rate increases, the individual household decision increases. In line with the above, we exploit the information channel between adopter and non-adopter households by taking advantage of the geographical coordinates through the Global Position System (GPS). The inclusion of a machine learning clustering algorithm allows to assess the impact of social interaction degree, by taking the physical distance into account, on the probability of adoption. The analysis of this empirical measure suggests that as the minimum distance in kilometers by road from a non-adopter to an adopter increases, the probability adoption rate decreases. This effect represents the most striking result since it captures the spillover and neighborhood effects widely recognized in literature. Our analysis also allows to deal with correlated non-observable or covariates omission at the village level. The evidence shows that the two variable of interest are robust to the additional controls. We also construct two additional measures of networks effects on the basis of social networks approach. The interest in computing them lies in stressing the role of simultaneity or reflection problem discussed in the empirical literature. The regression analysis on it points out that the closeness and betweenness centrality determines the effects on the probability of adoption in the right direction but results statistically non-significance.

Future research on this field must look into more detailed data sets that can provide aggregate and individual networks variables, allowing for more evidence on social interactions than only using geographical patterns. This is because the degree of interaction is also captured thorough the family and community ties and other kind of relationships which are not considered in our data set. In this sense, new developments on this branch of research in agricultural economics must open up new opportunities to cover additional analysis on other types of crops.

References

- [1] Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.
- [2] Acheampong, P. and Owusu, V. (2015). Impact of improved cassava varieties’ adoption on farmers’ incomes in rural ghana. 2015 Conference, August 9-14, 2015, Milan, Italy 210875, International Association of Agricultural Economists.
- [3] Afolami, A., Obayelu, E., and Vaughan, I. (2015). Welfare impact of adoption of improved cassava varieties by rural households in south western nigeria. *Agricultural and Food Economics*, pages 3–18.
- [4] Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica*, 65(5):1005–1027.
- [5] Baerenklau, K. A. (2005). Toward an understanding of technology adoption: Risk, learning, and neighborhood effects. *Land Economics*, 81(1):1–19.
- [6] Bandiera, O. and Rasul, I. (2002). Social networks and technology adoption in northern mozambique. *Working Paper*, pages 1–37.
- [7] Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.
- [8] Beaman, L., Yishay, A., Magruder, J., and Mobarak, A. (2015). Can network theory-based targeting increase technology adoption? *Working Paper*, pages 1–64.
- [9] Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.
- [10] Besley, T. and Case, A. (1993). Modeling technology adoption in developing countries. *The American Economic Review*, 83(2):396–402.
- [11] Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *The American Economic Review*, 85(3):365–390.
- [12] Brock, W. A. and Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260.
- [13] Brock, W. A. and Durlauf, S. N. (2003). Multinomial choice with social interactions. Working Paper 288, National Bureau of Economic Research.

- [14] Brock, W. A. and Durlauf, S. N. (2007). Discrete choice with social interactions and endogenous memberships. *Journal of Econometrics*, 140(1):52–75.
- [15] Cameron, L. A. (1999). The importance of learning in the adoption of high-yielding variety seeds. *American Journal of Agricultural Economics*, 81(1):83–94.
- [16] Carvalho, V. and Voigtländer, N. (2014). Input diffusion and the evolution of production networks. Technical report.
- [17] Coleman, J., Campbell, E., Hobson, J., McPartland, J., Mood, A., Weinfeld, F., and York, R. (1966). *Equality of Education Opportunity*.
- [18] Conley, T. G. and Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, 17(4):303–327.
- [19] Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review*, 100(1):35–69.
- [20] Cooper, R. and John, A. (1988). Coordinating coordination failures in keynesian models. *The Quarterly Journal of Economics*, 103(3):441–463.
- [21] Feder, G., Just, R. E., and Zilberman, D. (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change*, 33(2):255–298.
- [22] Foster, A. D. and Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6):1176–1209.
- [23] Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–548.
- [24] Goolsbee, A. and Klenow, P. (2002). Evidence on learning and network externalities in the diffusion of home computers. *The Journal of Law & Economics*, 45(2):317–343.
- [25] Gowrisankaran, G. and Stavins, J. (2004). Network externalities and technology adoption: Lessons from electronic payments. *The RAND Journal of Economics*, 35(2):260–276.
- [26] Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4):501–522.
- [27] Howeler, R. H., Lutaladio, N., Thomas, G., Food, and of the United Nations., A. O. (2013 2013). *Save and grow : cassava : a guide to sustainable production intensification / prepared by Reinhardt Howeler, NeBambi Lutaladio and Graeme Thomas*. Food and Agriculture Organization of the United Nations Rome.
- [28] Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press, Princeton, NJ, USA.
- [29] Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, (22):3–42.

-
- [30] Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.
 - [31] Manski, C. F. (2000). Economic analysis of social interactions. *The Journal of Economic Perspectives*, 14(3):115–136.
 - [32] Marini, A. (2011). Cultural beliefs and individual entity: Estimating a social interactions model on us data. *Working Paper*, pages 1–37.
 - [33] Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5):1002–1037.
 - [34] Woittiez, I. and Kapteyn, A. (1998). Social interactions and habit formation in a model of female labor supply. *Journal of Public Economics*, 68(2):185–205.
 - [35] Yishay, A. and Mobarak, A. (2015). Social learning and incentives for experimentation and communication. *Working Paper*, pages 1–45.
 - [36] Zanella, G. (2007). Discrete choice with social interactions and endogenous memberships. *Journal of the European Economic Association*, 5(1):122–153.
 - [37] Zeddies, J., Schaab, R., and Neuenschwander, P. (2001). Economics of biological control of cassava mealybug in africa. *Agricultural Economics*, 24(2):209–219.

A. Graphs and Tables

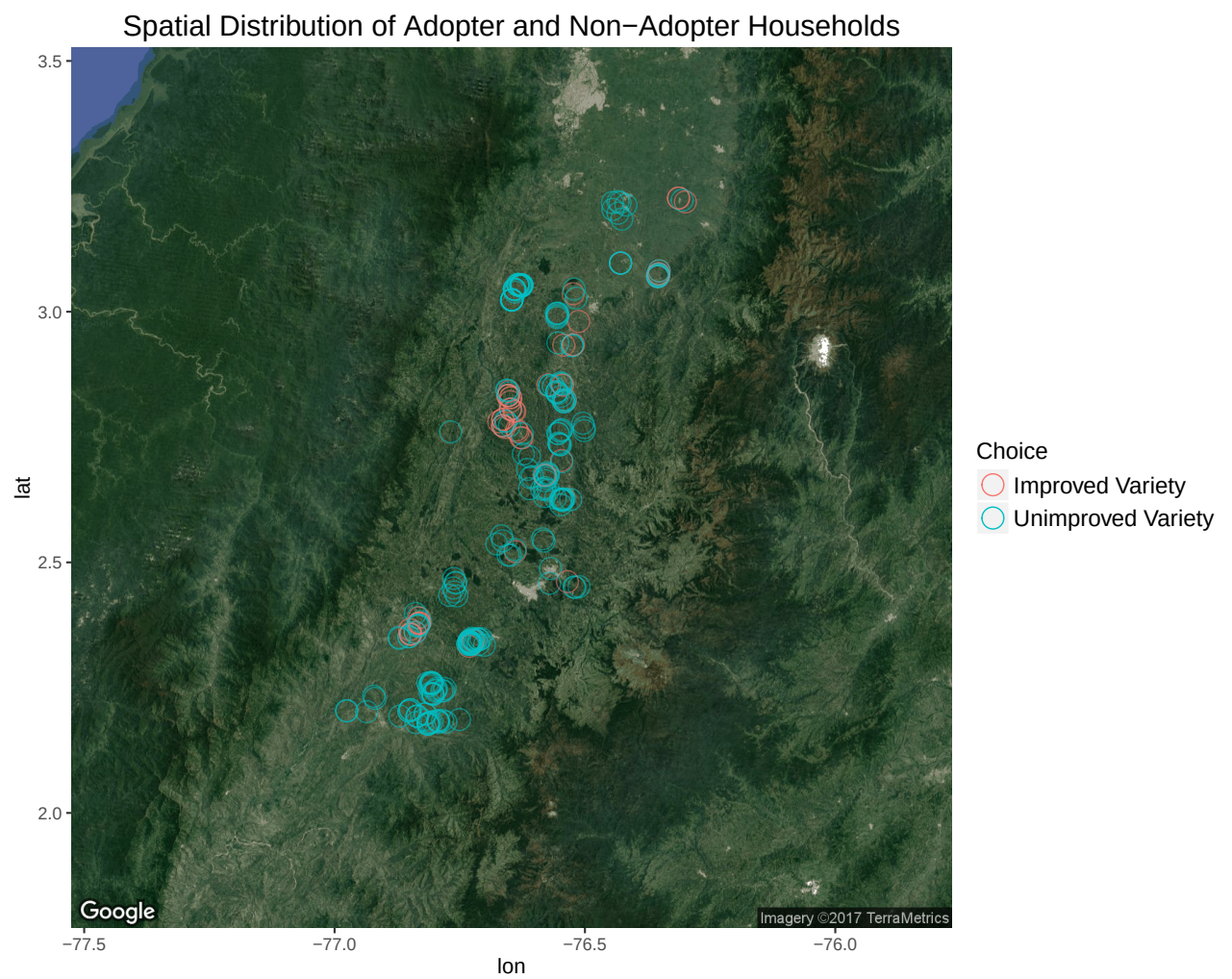


Fig. A.1 Source: Author calculations.



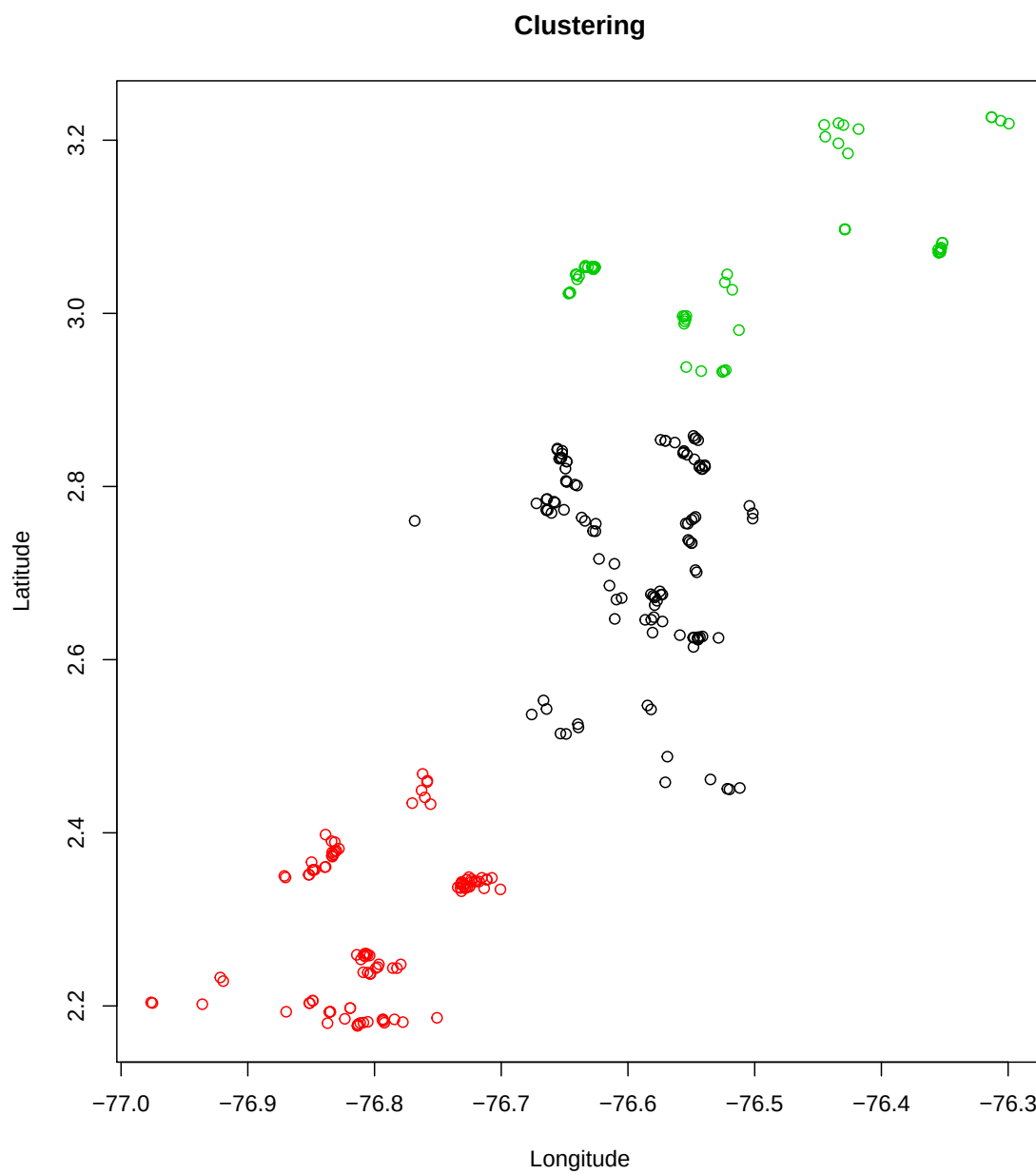


Fig. A.3 Source: Author calculations.

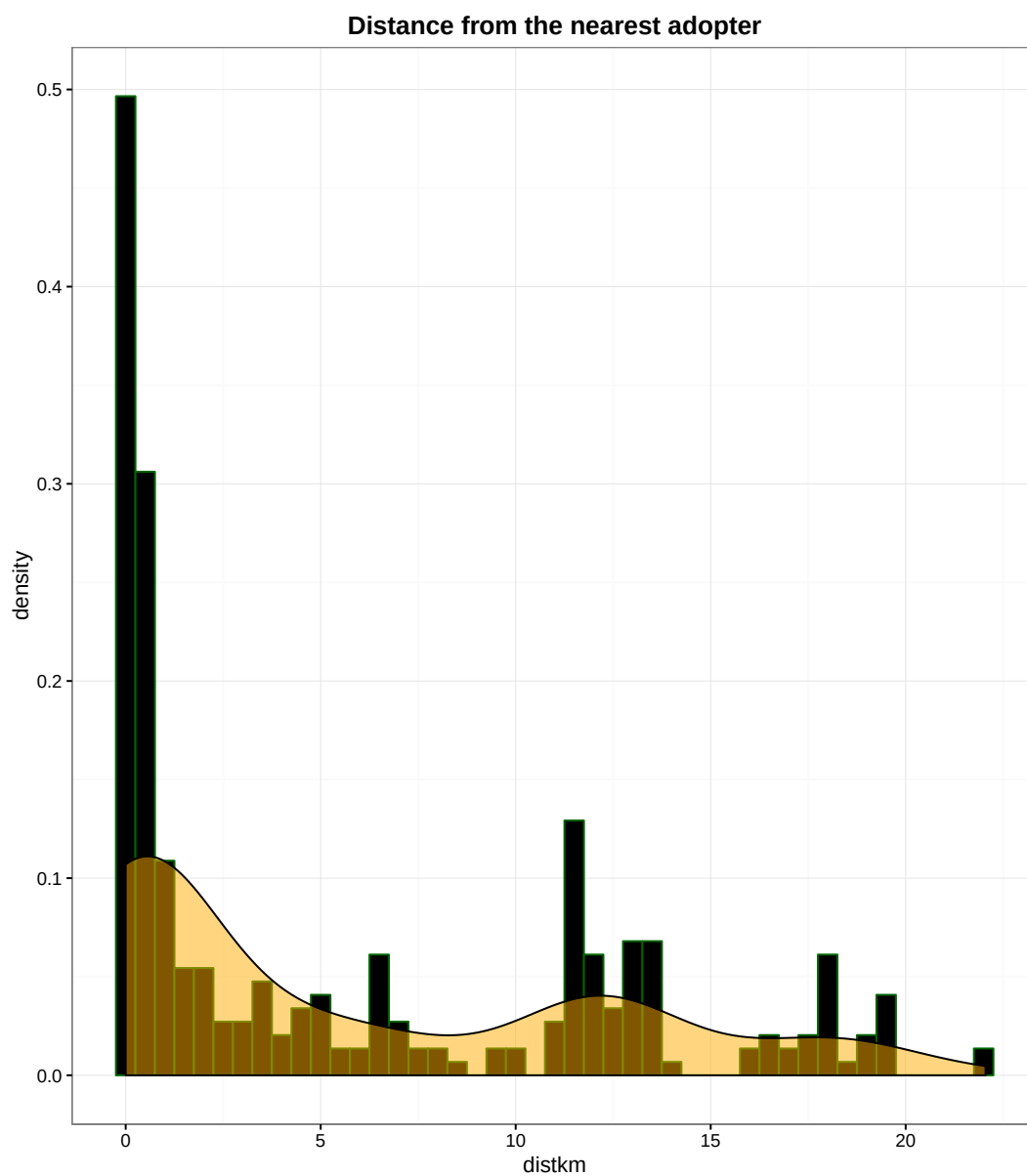


Fig. A.4 Source: Author calculations.

Table A.1 Variables Description

Variable name	Definition
Adoption	Dummy (1 if household <i>i</i> adopted an improved variety)
Distance from the nearest adopter	Distance in kilometers by road from the nearest adopter
Height above sea level	Height in meters above sea level
Farm size	Area in hectares of the farm
Cassava area	Area in hectares of the crop
Number of Appliances	Total number of appliances a household owns
Head of household's scholaryity	Education level in years
Head of household's age	Age in years
Household size	Total number of household members
Credit acces	Dummy (1 if household <i>i</i> got financial resources)
Technical acces	Dummy (1 if household <i>i</i> got training talks on e. s)
Weed control	Dummy (1 if household <i>i</i> applied weed controls)
Fertilizer	Dummy (1 if household <i>i</i> applied fertilizers)
Years of cassava crop experience	Years of experience
Number of members who work	Dummy (1 only if one households member worked)
Low social stratum	Dummy (1 if household belonged to 0,1,2 stratum)
Climate change	Dummy (1 if household <i>i</i> believed in high risk of climate)
Housing Quality	Score assigned by DANE (2005)
Phone Coverage	Percentage of phone coverage, DANE (2005)
Human Capital	Score assigned by DANE for the rural area (2005)
Unsatisfied Basic Needs	Percentage of people for the rural area, DANE (2005)
Quality Life Index	Score assigned by DNP (2005)

Table A.2 Summary Statistics

Variable	N	Mean	St. Dev.	Min	Max
Adoption	293	0.177	0.383	0	1
Height above sea level	293	1,452.089	316.819	885	2,078
Farm size	293	3.063	3.848	0.160	40.000
Cassava area	293	1.334	2.065	0.000	22.000
Number of appliances	293	13.997	5.166	0	49
Head of household's scholarship	293	5.163	3.441	0.000	21.000
Head of household's age	293	52.208	14.257	23	91
Household size	293	3.887	1.837	1	11
Credit access	293	0.491	0.501	0	1
Technical access	293	0.072	0.258	0	1
Weed control	293	0.823	0.383	0	1
Fertilizer	293	0.625	0.485	0	1
Years of cassava crop experience	293	19.341	13.241	1	60
Number of members who work (1)	293	0.628	0.484	0	1
Number of members who work (>1)	293	0.331	0.471	0	1
Low social stratum	293	0.990	0.101	0	1
Intermediate social stratum	293	0.003	0.058	0	1
Climate change	293	0.891	0.312	0	1
Housing Quality	293	5.815	1.415	3.860	9.970
Phone Coverage	293	10.219	13.926	1.570	70.610
Human Capital	293	22.202	1.749	19.870	28.430
Unsatisfied Basic Needs	293	56.349	11.881	14.560	71.820
Quality Life Index	293	58.468	10.429	48.010	84.400

Table A.3 Adoption Rate by Village

Village	Village Adoption Rate
Buenos Aires	0.00
Cajibío	0.00
Caldono	0.14
Caloto	0.29
El Bordo	0.00
El Tambo	0.33
La Sierra	0.00
Morales	0.60
Padilla	0.75
Piendamó	0.08
Popayán	0.20
Puerto Tejada	0.00
Rosas	0.00
Santander de Quilichao	0.18
Timbío	0.08
Villa Rica	0.00

Table A.4 Regression Results I

	Technology Adoption Regressions			
	Adoption (= 1)			
	(1)	(2)	(3)	(4)
Village adoption rate	0.980*** (0.070)	0.954*** (0.098)	0.903*** (0.103)	
Height above sea level (Meters)		−0.0001 (0.0001)	−0.0001* (0.0001)	−0.0003*** (0.0001)
Farm size (Hectare)		0.013* (0.007)	0.013* (0.007)	0.013* (0.007)
Cassava area (Hectare)		−0.006 (0.012)	−0.001 (0.013)	0.002 (0.013)
Number of appliances		−0.009** (0.004)	−0.008* (0.004)	−0.007 (0.004)
Head of household's scholarship		0.014** (0.006)	0.017*** (0.006)	0.017*** (0.006)
Head of household's age		−0.001 (0.001)	−0.00004 (0.002)	−0.0001 (0.001)
Household size		−0.004 (0.011)	−0.006 (0.012)	−0.010 (0.011)
Credit access (indicator variable)		0.026 (0.039)	0.026 (0.040)	0.012 (0.041)
Technical access (indicator variable)		0.301*** (0.073)	0.286*** (0.074)	0.308*** (0.076)
Weed control (indicator variable)		0.085* (0.049)	0.094* (0.050)	0.104** (0.051)
Fertilizer (indicator variable)		−0.019 (0.041)	−0.012 (0.042)	−0.014 (0.047)
Years of cassava crop experience		0.004** (0.002)	0.003** (0.002)	0.003** (0.001)
Number of members who work (One-Indicator variable)		0.113 (0.089)	0.172* (0.095)	0.191* (0.095)
Number of members who work (More than one-indicator variable)		0.078 (0.095)	0.138 (0.100)	0.167 (0.101)
Low social stratum (indicator variable)		−0.018 (0.138)	0.312 (0.245)	0.403 (0.269)
Intermediate social stratum (indicator variable)		−0.552 (0.360)	−0.258 (0.401)	−0.167 (0.423)
Climate change		−0.018 (0.060)	−0.010 (0.061)	−0.021 (0.061)
Housing quality			0.036 (0.042)	0.373 (0.334)
Phone coverage			0.001 (0.003)	0.016 (0.034)
Human capital			−0.033 (0.029)	−0.096* (0.046)
Unsatisfied Basic Needs			0.001 (0.004)	0.025 (0.022)
Quality Life Index			0.0004 (0.009)	−0.026 (0.055)
Fixed Effects	NO	NO	NO	YES
Observations	293	293	293	293
R ²	0.414	0.499	0.506	0.521
Adjusted R ²	0.412	0.466	0.464	0.460

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.5 Regression Results II

	Technology Adoption Regressions			
	Adoption (= 1)			
	(1)	(2)	(3)	(4)
Distance in KM from the nearest adopter	0.004 (0.003)	−0.026*** (0.004)	−0.030*** (0.004)	
Height above sea level (Meters)		−0.0002*** (0.0001)	−0.0002** (0.0001)	−0.0003 (0.0001)
Farm size (Hectare)		0.016** (0.007)	0.014** (0.007)	0.013* (0.007)
Cassava area (Hectare)		0.012 (0.013)	0.014 (0.013)	0.002 (0.013)
Number of appliances		−0.008* (0.005)	−0.006 (0.005)	−0.007* (0.004)
Head of household's scholarship		0.011* (0.006)	0.015** (0.006)	0.017*** (0.006)
Head of household's age		−0.001 (0.002)	−0.00003 (0.002)	−0.0001 (0.001)
Household size		−0.026** (0.012)	−0.023* (0.012)	−0.010 (0.011)
Credit access (indicator variable)		0.001 (0.042)	0.007 (0.042)	0.012 (0.041)
Technical access (indicator variable)		0.378*** (0.077)	0.342*** (0.078)	0.308*** (0.076)
Weed control (indicator variable)		0.130** (0.052)	0.123** (0.052)	0.104** (0.051)
Fertilizer (indicator variable)		−0.052 (0.044)	−0.035 (0.045)	−0.014 (0.047)
Years of cassava crop experience		0.002 (0.002)	0.003 (0.002)	0.003** (0.001)
Number of members who work (One-indicator variable)		0.183* (0.095)	0.207** (0.099)	0.191* (0.095)
Number of members who work (More than one-indicator variable)		0.174* (0.101)	0.167* (0.104)	0.167 (0.101)
Low social stratum (indicator variable)		0.468*** (0.153)	0.827*** (0.247)	0.403 (0.269)
Intermediate social stratum (indicator variable)		0.249 (0.380)	0.467 (0.409)	−0.167 (0.423)
Climate change (indicator variable)		−0.037 (0.064)	−0.014 (0.064)	−0.021 (0.061)
Housing quality			−0.113** (0.051)	0.373 (0.334)
Phone coverage			0.005 (0.004)	0.016 (0.034)
Human capital			0.072** (0.034)	−0.096* (0.046)
Unsatisfied Basic Needs Index			−0.010** (0.004)	0.025 (0.022)
Quality Life Index			−0.016* (0.009)	−0.026 (0.055)
Fixed Effects	NO	NO	NO	YES
Observations	293	293	293	293
R ²	0.005	0.431	0.459	0.521
Adjusted R ²	0.002	0.394	0.413	0.460

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.6 Regression Results III

	Technology Adoption Regressions	
	Adoption (= 1)	
	(1)	(2)
Distance in KM from the nearest adopter (driving mode)	–0.009*** (0.001)	
Time in minutes from the nearest adopter (driving mode)		–0.004*** (0.001)
Height above sea level (Meters)	–0.0001 (0.0001)	–0.0001* (0.0001)
Farm size (Hectare)	0.015** (0.007)	0.015** (0.007)
Cassava area (Hectare)	0.014 (0.013)	0.014 (0.013)
Number of appliances	–0.006 (0.005)	–0.004 (0.005)
Head of household's scholarship	0.014** (0.006)	0.014** (0.007)
Head of household's age	0.0001 (0.002)	0.0003 (0.002)
Household size	–0.024** (0.012)	–0.028** (0.012)
Credit access (indicator variable)	0.017 (0.042)	0.016 (0.042)
Technical access (indicator variable)	0.312*** (0.078)	0.308*** (0.078)
Weed control (indicator variable)	0.118** (0.052)	0.119** (0.052)
Fertilizer (indicator variable)	–0.034 (0.045)	–0.029 (0.045)
Years of cassava crop experience	0.002 (0.002)	0.002 (0.002)
Number of members who work (One-Indicator variable)	0.176* (0.101)	0.177* (0.101)
Number of members who work (More than one-indicator variable)	0.149 (0.106)	0.152 (0.106)
Low social stratum (indicator variable)	0.810*** (0.249)	0.838*** (0.250)
Intermediate social stratum (indicator variable)	0.465 (0.412)	0.501 (0.413)
Climate change (indicator variable)	–0.006 (0.064)	–0.006 (0.064)
Housing quality	–0.064 (0.049)	–0.025 (0.047)
Phone coverage	0.003 (0.004)	0.004 (0.004)
Human capital	0.024 (0.032)	0.003 (0.031)
Unsatisfied Basic Needs	–0.005 (0.004)	–0.003 (0.004)
Quality Life Index	–0.009 (0.009)	–0.008 (0.009)
Observations	293	293
R ²	0.450	0.446
Adjusted R ²	0.403	0.399

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.7 Regression Results IV

	Technology Adoption Regressions				
	Adoption (= 1)				
	(1)	(2)	(3)	(4)	(5)
Distance in KM from the nearest adopter	−0.032*** (0.004)				
Distance in KM from the nearest trained adopter		−0.003** (0.001)			
Distance in KM from the nearest adopter in weed control			−0.032*** (0.004)		
Distance in KM from the nearest adopter in fertilizer				−0.030*** (0.004)	
Imitation effect					0.175 (0.143)
Height above sea level (Meters)	−0.0001 (0.0001)	−0.0002** (0.0001)	−0.0001 (0.0001)	−0.0001 (0.0001)	−0.0001* (0.0001)
Farm size (Hectare)	0.015** (0.007)	0.015* (0.008)	0.015** (0.007)	0.014** (0.007)	0.013* (0.008)
Cassava area (Hectare)	0.011 (0.013)	0.018 (0.014)	0.011 (0.013)	0.011 (0.013)	0.018 (0.014)
Number of appliances	−0.007 (0.004)	−0.002 (0.005)	−0.007 (0.004)	−0.007 (0.004)	−0.003 (0.005)
Head of household's scholarship	0.013** (0.006)	0.018*** (0.007)	0.013** (0.006)	0.014** (0.006)	0.020*** (0.007)
Head of household's age	−0.001 (0.002)	0.001 (0.002)	−0.0005 (0.002)	−0.0002 (0.002)	0.001 (0.002)
Household size	−0.024** (0.011)	−0.027** (0.013)	−0.023** (0.011)	−0.024** (0.012)	−0.029** (0.013)
Credit access	0.004 (0.040)	0.031 (0.045)	0.005 (0.040)	0.009 (0.040)	0.032 (0.045)
Technical access (indicator variable)	0.315*** (0.074)	0.292*** (0.084)	0.314*** (0.074)	0.305*** (0.075)	0.328*** (0.084)
Weed control (indicator variable)	0.123** (0.050)	0.115** (0.056)	0.121** (0.050)	0.123** (0.050)	0.122** (0.056)
Fertilizer (indicator variable)	−0.012 (0.042)	0.009 (0.049)	−0.014 (0.042)	−0.025 (0.043)	0.036 (0.047)
Years of cassava crop experience	0.003* (0.002)	0.002 (0.002)	0.003* (0.002)	0.003* (0.002)	0.001 (0.002)
Number of members who work (One-Indicator variable)	0.190** (0.095)	0.248** (0.106)	0.188** (0.095)	0.212** (0.096)	0.258** (0.107)
Number of members who work (More than one-indicator variable)	0.185* (0.100)	0.205* (0.112)	0.183* (0.100)	0.203** (0.101)	0.226** (0.112)
Low social stratum (indicator variable)	0.817*** (0.237)	0.919*** (0.266)	0.816*** (0.237)	0.866*** (0.240)	0.865*** (0.267)
Intermediate social stratum (indicator variable)	0.462 (0.392)	0.579 (0.439)	0.458 (0.392)	0.497 (0.397)	0.551 (0.441)
Climate change	−0.018 (0.061)	0.001 (0.068)	−0.016 (0.061)	−0.020 (0.062)	0.006 (0.069)
Housing quality	−0.141*** (0.049)	0.035 (0.050)	−0.142*** (0.049)	−0.126** (0.049)	0.076 (0.048)
Phone coverage	0.005 (0.003)	0.005 (0.004)	0.004 (0.003)	0.008** (0.003)	0.007* (0.004)
Human capital	0.090*** (0.033)	−0.016 (0.034)	0.088*** (0.032)	0.091*** (0.033)	−0.035 (0.033)
Unsatisfied Basic Needs	−0.013*** (0.004)	−0.002 (0.004)	−0.013*** (0.004)	−0.013*** (0.004)	0.001 (0.004)
Quality Life Index	−0.018** (0.008)	−0.012 (0.009)	−0.017** (0.008)	−0.023*** (0.009)	−0.014 (0.010)
Observations	293	293	293	293	293
R ²	0.501	0.376	0.501	0.490	0.369
Adjusted R ²	0.458	0.323	0.459	0.446	0.315

Note: No fixed effects at village level

*p<0.1; **p<0.05; ***p<0.01

Table A.8 Regression Results V

	Technology Adoption Regressions			
	Adoption (= 1)			
	(1)	(2)	(3)	(4)
Closeness	−0.019*** (0.002)	0.109*** (0.036)	−0.100 (0.110)	
Height above sea level (Meters)		−0.0001 (0.0001)	−0.0002** (0.0001)	−0.0003*** (0.0001)
Farm size (Hectare)		0.015** (0.008)	0.012 (0.008)	0.013* (0.007)
Cassava area (Hectare)		0.021 (0.014)	0.017 (0.014)	0.002 (0.013)
Number of appliances		−0.003 (0.005)	−0.004 (0.005)	−0.007 (0.004)
Head of household's scholarship		0.016** (0.007)	0.020*** (0.007)	0.017*** (0.006)
Head of household's age		0.001 (0.002)	0.001 (0.002)	−0.0001 (0.001)
Household size		−0.027** (0.013)	−0.027** (0.013)	−0.010 (0.011)
Credit access (indicator variable)		0.019 (0.044)	0.039 (0.045)	0.012 (0.041)
Technical access (indicator variable)		0.335*** (0.083)	0.318*** (0.084)	0.308*** (0.076)
Weed control (indicator variable)		0.129** (0.056)	0.118** (0.056)	0.104** (0.051)
Fertilizer (indicator variable)		0.026 (0.046)	0.055 (0.049)	−0.014 (0.047)
Years of cassava crop experience		0.001 (0.002)	0.001 (0.002)	0.003** (0.001)
Number of members who work (One-Indicator variable)		0.241** (0.107)	0.263** (0.107)	0.191* (0.095)
Number of members who work (More than one-indicator variable)		0.203* (0.113)	0.224** (0.113)	0.167 (0.101)
Low social stratum (indicator variable)		0.888*** (0.291)	0.757** (0.297)	0.403 (0.269)
Intermediate social stratum (indicator variable)		0.633 (0.457)	0.431 (0.463)	−0.167 (0.423)
Climate change		−0.006 (0.069)	0.007 (0.069)	−0.021 (0.061)
Housing quality			0.065 (0.048)	0.373 (0.334)
Phone coverage			0.007* (0.004)	0.016 (0.034)
Human capital			−0.070 (0.047)	−0.096* (0.046)
Unsatisfied Basic Needs			−0.002 (0.005)	0.025 (0.022)
Quality Life Index			−0.010 (0.010)	−0.026 (0.055)
Fixed Effects	NO	NO	NO	YES
Observations	293	293	293	293
R ²	0.176	0.347	0.367	0.521
Adjusted R ²	0.173	0.304	0.314	0.460

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.9 Regression Results VI

	Technology Adoption Regressions			
	Adoption (= 1)			
	(1)	(2)	(3)	(4)
Betweenness	0.102** (0.040)	0.009 (0.037)	0.006 (0.036)	
Height above sea level (Meters)		−0.0001 (0.0001)	−0.0001* (0.0001)	−0.0003*** (0.0001)
Farm size (Hectare)		0.013* (0.008)	0.013* (0.008)	0.013* (0.007)
Cassava area (Hectare)		0.011 (0.014)	0.019 (0.014)	0.002 (0.013)
Number of appliances		−0.006 (0.005)	−0.003 (0.005)	−0.007 (0.004)
Head of household's scholarship		0.012* (0.007)	0.020*** (0.007)	0.017*** (0.006)
Head of household's age		−0.001 (0.002)	0.001 (0.002)	−0.0001 (0.001)
Household size		−0.027** (0.013)	−0.028** (0.013)	−0.010 (0.011)
Credit access (indicator variable)		0.025 (0.045)	0.033 (0.045)	0.012 (0.041)
Technical access (indicator variable)		0.372*** (0.084)	0.320*** (0.084)	0.308*** (0.076)
Weed control (indicator variable)		0.113** (0.057)	0.124** (0.056)	0.104** (0.051)
Fertilizer (indicator variable)		0.023 (0.047)	0.043 (0.047)	−0.014 (0.047)
Years of cassava crop experience		0.001 (0.002)	0.001 (0.002)	0.003** (0.001)
Number of members who work (One-Indicator variable)		0.144 (0.104)	0.268** (0.107)	0.191* (0.095)
Number of members who work (More than one-indicator variable)		0.105 (0.110)	0.231** (0.113)	0.167 (0.101)
Low social stratum (indicator variable)		0.146 (0.159)	0.875*** (0.268)	0.403 (0.269)
Intermediate social stratum (indicator variable)		−0.001 (0.412)	0.559 (0.442)	−0.167 (0.423)
Climate change		−0.024 (0.070)	0.002 (0.069)	−0.021 (0.061)
Housing quality			0.070 (0.048)	0.373 (0.334)
Phone coverage			0.006 (0.004)	0.016 (0.034)
Human capital			−0.039 (0.033)	−0.096* (0.046)
Unsatisfied Basic Needs			0.001 (0.004)	0.025 (0.022)
Quality Life Index			−0.011 (0.010)	−0.026 (0.055)
Fixed Effects	NO	NO	NO	YES
Observations	293	293	293	293
R ²	0.022	0.325	0.366	0.521
Adjusted R ²	0.018	0.281	0.312	0.460

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.10 Logistic Regression Results

	Technology Adoption Regressions	
	Adoption (= 1)	
	(1)	(2)
Village adoption rate	8.941*** (1.618)	
Distance in KM from the nearest adopter		−0.689*** (0.161)
Height above sea level (Meters)	−0.002** (0.001)	−0.001 (0.001)
Farm size (Hectare)	0.097 (0.066)	0.105 (0.064)
Cassava area (Hectare)	−0.012 (0.135)	0.021 (0.136)
Number of appliances	−0.110* (0.060)	−0.108* (0.058)
Head of household's scholarship	0.141** (0.068)	0.113* (0.067)
Head of household's age	−0.015 (0.020)	−0.010 (0.019)
Household size	−0.141 (0.136)	−0.145 (0.126)
Credit access (indicator variable)	0.343 (0.451)	0.222 (0.450)
Technical access (indicator variable)	2.641*** (0.784)	3.781*** (0.986)
Weed control (indicator variable)	1.372* (0.803)	2.441*** (0.892)
Fertilizer (indicator variable)	0.124 (0.536)	−0.887 (0.563)
Years of cassava crop experience	0.048** (0.020)	0.055*** (0.020)
Climate change (indicator variable)	−0.152 (0.717)	−0.376 (0.816)
Housing Quality	2.075** (1.011)	−2.106** (0.987)
Phone Coverage	0.006 (0.037)	0.024 (0.038)
Human Capital	−1.290** (0.652)	1.245** (0.626)
Unsatisfied Basic Needs	0.142* (0.077)	−0.126* (0.076)
Quality Life Index	0.103 (0.111)	−0.130 (0.116)
Observations	293	293
Log Likelihood	−78.470	−75.965
Akaike Inf. Crit.	194.940	189.931

Note: Marginal effects are not reported

*p<0.1; **p<0.05; ***p<0.01