CAM
Centre for Applied Microeconometrics

Department of Economics
University of Copenhagen

http://www.econ.ku.dk/CAM/

Working paper no. 2015-01

The network at work: Diffusion of banana cultivation in Tanzania

Anna Folke Larsen

The network at work:

Diffusion of banana cultivation in Tanzania

Anna Folke Larsen*

11th May 2015

Abstract

This paper investigates the role of networks for diffusion of improved banana cultivation introduced by an agricultural project in Tanzania. In the existing literature on networks and technology adoption, network effects are interpreted as learning. I show that a farmer's network can affect the adoption of a new crop not only through social learning, but also by providing necessary inputs for adoption. I set up a simple model for adoption and derive similar model implications for the provision of either inputs or information through the network. Empirically, I find that a farmer is 37 percentage points more likely to adopt banana cultivation if there is at least one project participant growing bananas in the farmer's network. I use three falsification tests to support causal interpretation of the network effect on adoption. Provision of inputs (banana seedlings) through networks is found to play an important role for the network effects found.

1 Introduction

There are huge disparities in agricultural productivity across countries with agricultural output per worker being more than 100 times larger in the United States than in Sub-Saharan African countries (Gollin et al., 2014). As the majority of poor people in developing countries are employed in the agricultural sector, agricultural growth has the strongest potential compared to other sectors to reduce poverty in developing countries, in particular among the poorest of the poor (Ligon & Sadoulet, 2011; de Janvry & Sadoulet, 2010; Christiaensen et al., 2011).[1] Though Africa's Green Revolution has been a long time coming, agricultural growth in Sub-Saharan African countries is still key to transforming their economies and reducing poverty. Indeed, population growth and declining farm sizes call for locally adapted technological change in the agricultural sector (Diao et al., 2010). Moreover, climate change increases the

*Department of Economics, University of Copenhagen, email: afl@econ.ku.dk, tel.: +4561661159. I would like to thank the Rockwool Foundation for financial support for the data collection, Helene Bie Lilleør, Mette Ejrnæs, Bo Honore, and Peter Norman Sørensen for invaluable comments and guidance, Maria Fibæk and Dominick Ringo, Catherine Maguzu and all other RECODA staff for great support during field work, Jens Vesterager for agricultural technical guidance and Cathrine Søgaard Hansen for research assistance. All errors remain my own.

[1]The contribution to poverty reduction from the agricultural sector stems not only from the size of the sector and the participation of poor people in the sector, but also from its indirect impact on growth in other sectors (Christiaensen et al., 2011).

necessity of technological change in agriculture to adapt to the more erratic rainfall (Lybbert & Sumner, 2012).

Hence, understanding barriers to adoption and diffusion of new agricultural technologies is key for agricultural development, poverty reduction and adaptation to climate change. The topic is not new; adoption of new agricultural technologies has been studied in a variety of countries and settings since the seminal work by Griliches (1957) (see reviews by Foster & Rosenzweig, 2010; Sunding & Zilberman, 2001; Evenson & Westphal, 1995; Feder et al., 1985).

In this paper I study how a farmer's network affects the decision to adopt a new agricultural technology in the context of African small-scale farming. The existing literature focus on the role of *social learning* through networks (Carter et al., 2014; Magnan et al., 2015; Conley & Udry, 2010; Bandiera & Rasul, 2006; Munshi, 2004; Foster & Rosenzweig, 1995; Krishnan & Patnam, 2014). These studies suggest that the network helps to relax an informational constraint faced by the farmer. I contribute to this literature by showing that a farmer's network can affect the adoption of a new crop not only through social learning, but also by providing necessary inputs for adoption. To my knowledge, the *provision of inputs* through networks has not been studied as an alternative or a complement to social learning.[2] This is an important distinction both because the role of information through networks may be exaggerated if it is confounded by input provision, but in particular because networks have the potential to mitigate not only imperfect information, but also input market imperfections. This can be used deliberately when designing future projects to increase diffusion of agricultural technologies, in particular in remote areas where input distribution is complicated by poor infrastructure.

Indeed, Spencer (1996) argues that the Green Revolution in Africa has been hindered by a low coverage of rural roads which impedes the distribution of inputs such as improved seeds and fertilizer. Road density in the low-income countries of Sub-Saharan Africa is less than half of that in low-income countries in the rest of the world (Carruthers et al., 2009). Suri (2011) confirms that malfunctioning input markets hamper adoption of hybrid maize in Kenya despite large gross returns. Access to improved pigeon pea seeds is also found to improve net gains remarkably in Tanzania (Shiferaw et al., 2008).

I study the adoption of improved banana cultivation in the Arusha region in Tanzania. Improved banana variety seedlings and a new banana cultivation technique were introduced to participants of a Farmer Field School project called RIPAT in eight villages. The RIPAT project was designed to foster diffusion of banana cultivation to non-participants through a solidarity chain principle: Participants where obliged to pass on thrice as many seedlings as they received through the project to other farmers, free of charge. As the improved banana variety seedlings were not available through formal channels, input provision through networks becomes very important for adoption in this context.

[2]Emerick (2013) study the efficiency of input provision through networks as opposed to door-to-door visit, but he does not consider information provision through networks. Besides, he does not have data on networks but rely on sub-caste and last name as proxies for network connections.

To guide intuition for the adoption behavior among non-participating farmers, I set up a simple model of crop choice. Following the literature on social learning, I first derive model implications under imperfect information and show that adopters in the network can affect the farmer's adoption decision through information about expected yields of the new crop. I then extend the model by allowing the network to provide inputs for the new crop when there is an imperfect input market. I can derive the exact same model implications with an imperfect input market as under imperfect information.

Turning to the data, I explore how the adoption among non-participants in the project villages depends on their informational links to project participants and to other farmers. I use data on 509 non-participating farmers from households within the eight RIPAT villages collected for the purpose of this study. I find large network effects on adoption behavior: discussing farming issues with at least one RIPAT farmer growing improved bananas increases the propensity to adopt by 37 percentage points. The data suggest that provision of inputs through the network contribute to this very strong network effect.

I further add to the literature by showing that network members who do not grow improved bananas have a negative effect on the propensity to adopt. The theoretical model provides the following intuition: Network members not growing bananas provide information or inputs that makes other crops more attractive, reducing the relative profitability of bananas. For a given amount of land, the farmer is then less likely to adopt banana cultivation. This finding points to the importance of controlling for network size when assessing the impact on adoption of adoption behavior in the network. Failing to do so (e.g. as in Bandiera & Rasul, 2006) potentially confounds the network estimate.

The network effects described are inherently difficult to identify (Manski, 1993, 2000; Brock & Durlauf, 2007; Bramoullé et al., 2009). Experimental variation in adoption in the network can facilitate identification (e.g. Magnan et al., 2015; Cai et al., 2015; Carter et al., 2014; Kremer & Miguel, 2007), but it is often not available. As participation in RIPAT is voluntary and hence subject to self-selection I must use a different approach to address causality.

First, I note that the network is captured *prior* to adoption (using a recall question), and there is a natural ordering in the timing of adoption as RIPAT participants are the first to be introduced to improved banana cultivation. This mitigates a concern of a simultaneity problem. Next, I carefully investigate the different confounding factors and perform three falsification tests to address whether the estimated network effects are confounded by a) contextual effects; b) correlated effects; and c) self-selection into RIPAT, where a) and b) refers to the terminology of Manski (1993). *Contextual effects* cover the impact of the characteristics (rather than the behavior) of network members on individual behavior. I exploit detailed data on the RIPAT farmers in the network to test if the network effects are driven by the socioeconomic characteristics of network members. I do not find the characteristics to be driving the network effects found. *Correlated effects* capture the correlation in behavior within the network which is due to a common environment or a correlation in unobserved characteristics. I capture local growing conditions by the number of adopters within a radius of a half kilometer of the household and by subvillage fixed effects. I control for previous or current cultivation of traditional bananas to capture unobserved preferences for banana cultivation or prior knowledge of banana cultivation. Furthermore, I address the potential correlation of unobservables within networks in a placebo study. The network measures cannot predict adoption of three placebo crops, which leads me to conclude that the network effects found are not driven by a correlation in openness to new crops within networks. Finally, self-selection into RIPAT creates a concern for the interpretation of the results. My interpretation of the network effect is that the farmer is affected by the *adoption behavior* in his or her network either through the information or input channel. But as participation in RIPAT is voluntary the non-participants in my sample have implicitly self-selected out of RIPAT. They may have chosen to do so because they have network members who participate and they expect to receive information and inputs from them. In that case, they have decided to adopt regardless of the adoption behavior in the network and I would expect them to adopt as soon as possible. Hence, I explore the difference between early and late adopters to test if this behavior is driving the results. I find that the strong network effects persist among late adopters supporting my interpretation of the network estimates. Taken together, none of the evidence suggests that the estimated network effects are confounded.

The remainder of the paper is structured as follows: Section 2 introduces the Farmer Field School project and the agricultural technology under study. In section 3 I set up a simple model of crop choice to illustrate how the adoption decision is affected by the network through either information or input provision and I derive testable implications of the model. I proceed with a description of the data and summary statistics in section 4, and subsequently, I present the empirical specification and estimation results in section 5. Section 6 discusses the identification of the network effects. In this section I address contextual and correlated effects and self-selection into RIPAT with three falsification tests and furthermore, I discuss how the provision of seedlings through networks could explain a part of the network effects found. Finally, section 7 concludes.

2 RIPAT and improved banana cultivation

The improved banana cultivation studied in this paper is introduced by a project called RIPAT (Rural Initiatives for Participatory Agricultural Transformation). RIPAT is a multifaceted agricultural and livestock project that aims to alleviate food insecurity and poverty among the participating households.[3] A series of RIPAT projects have been implemented, and this study considers the first RIPAT project which took place in eight villages in Arumeru district in Northern Tanzania from 2006 to 2009. It was implemented by a local NGO RECODA and funded by the Rockwool Foundation.

Two Farmer Field School (FFS) groups are established in each village consisting of 30-35 farmers each.

[3] See thorough information on the project at www.RIPAT.org.

The farmers sign up voluntarily, but are only considered if they are dealing with agriculture already and if they have between one and five acres of land (however not rigorously abided).[4] The FFS group cultivates a common plot, where RECODA facilitators demonstrate new agricultural techniques from a 'basket of options'. After learning about the new techniques and improved varieties the participating farmers can choose to adopt on their own farm the components that best fit their soil, water accessibility, availability of household labor and land, preferences and taste.

The main component in the basket of options (and the most successful in terms of adoption) is a new technique of banana cultivation which is studied in this paper. It consists of special instructions for how to prepare the hard-pan soil and establish and tend a banana plantation, in conjunction with the introduction of five improved banana varieties which are more drought resistant than the traditional bananas grown in the area. The preparation of the soil consists of digging a one cubic meter hole which is then filled with a mixture of top soil and farm yard manure before planting the improved banana seedling. The soil around the plant can thereby contain more moisture which makes the plant more drought tolerant. The improved banana cultivation facilitates large scale plantations which is not possible with the traditional techniques in this area. The banana cultivation technique is indeed new in the area; when RIPAT was introduced at village meetings some people would laugh when banana plantations were mentioned because they knew it was not feasible–at least not with the existing techniques. The other components of the project include conservation agriculture, crop diversification, improved animal husbandry, fruit and multipurpose trees, soil and water conservation and post-harvesting technologies.

The project is designed to facilitate dissemination of the introduced technologies and varieties in several ways. A *solidarity chain* is established where participating farmers are obliged to pass on thrice as many improved banana seedlings to other farmers as they have received, free of charge. In addition, Super Farmers are chosen among the RIPAT farmers and educated to teach other farmers about banana cultivation. They are selected by the groups themselves among the best farmers to practice and teach the new methods. Furthermore, two criteria were set up for the formation of the Farmer Field School groups to foster dissemination of technologies: First, only farmers who were socially acceptable people and willing to share with others were admitted into the groups. Second, since each of the villages consist of two to five subvillages which are not necessarily contiguous, it was ensured that all subvillages where represented in the FFS groups.

I study the role of networks in the local diffusion of improved banana cultivation from participating to non-participating farmers within the project villages. In particular, I study how the adoption of improved banana cultivation among non-RIPAT farmers residing in the project villages depend on whether they discuss farming issues with RIPAT farmers who have adopted the new technique. For expositional purposes I will only use the term 'improved' when it is important to distinguish between the existing

[4] The fact that RIPAT farmers self-select into the project also implies that the non-RIPAT farmers are a selected group. I discuss this issue in section 6.3.

'traditional' banana cultivation and the new technique. Henceforth, 'banana cultivation' refers to the new technique.

The solidarity chain principle for improved banana seedlings was important for the diffusion of improved banana cultivation as the seedlings could not be purchased through formal channels in the area. Once the banana plant is established it produces seedlings which the farmer can only use if he or she wants to expand the banana plantation and hence the opportunity cost of giving them away is low. This is different from annual crops where the opportunity costs of the seed corn is to eat it or plant it on your own farm as you have to replant the crop every year.

The solidarity chain reduces the investment costs related to the establishment of a banana plantation. However, the opportunity costs of land and labor may still be considerable. The labor investment related to the establishment of the plantation is large as it is a very strenuous task to dig the big holes in the hard soil and some farmers may even choose to hire casual labor to dig the holes at a rate of around 2,000 Tanzanian Shillings (1.25 US dollars) per hole.[5] Nevertheless, planting one or two banana plants is manageable and affordable for most farmers and a gradual expansion of the banana plantation can then be decided upon after testing the banana plantation on a small scale.

Figure 1 illustrates the adoption of banana cultivation among RIPAT farmers and the diffusion to non-RIPAT farmers over time. The maps are based on household GPS location and adoption information from the data presented in section 4.1.[6] Before project implementation in 2005 very few households in the sample had adopted banana cultivation. Already at the end of 2006, the first year of RIPAT, we see widespread adoption of banana cultivation among RIPAT households, and some few non-RIPAT households have followed suit at this early stage. By 2008 the number of adopting RIPAT households has almost doubled and the new technique is also catching on among non-RIPAT households. One year after the end of the project, 69 percent of the RIPAT households are growing improved bananas on their farm and the improved banana cultivation has spread to 20 percent of the non-RIPAT households. This is considered a very large degree of diffusion compared to the existing Farmer Field School literature, where only limited diffusion of the new technologies is documented (See reviews in Davis et al., 2012; Waddington et al., 2014). The high degree of adoption of improved banana cultivation among RIPAT participants and diffusion to non-RIPAT households suggests that banana cultivation indeed suits the local needs and preferences, that it is trialable at a smaller scale, and that it is profitable compared to existing crops and technologies.

[5] As noted by the anthropologist, Quentin Gausset, during field studies in the RIPAT villages.
[6] Data are collected in January 2011 and time of adoption is based on a recall question

3 A simple model of crop choice

To guide intuition for the empirical results, I set up a simple model that illustrates how the crop adoption decision is affected by the network of the farmer when the information is not perfect or when the input market is not functioning. The model allows me to derive testable implications for how the adoption decision is affected by the network of the farmer either through information or input provision.

I model how the adoption decision depends on the egocentric network including links to three different types of farmers: RIPAT farmers, non-RIPAT banana growers and other farmers who do not grow bananas. There are two main differences between the model I present and existing learning models (e.g. Foster & Rosenzweig, 1995; Bardhan & Udry, 1999; Munshi, 2004; Conley & Udry, 2010; Besley & Case, 1993; Banerjee, 1992; Saha et al., 1994): 1) I show how social learning and provision of inputs through the network can lead to the exact same network effects on adoption behavior; and 2) I allow for network members growing other crops than the main crop studied to affect the adoption decision. To my knowledge, this has not been done before.

I will focus on model predictions for the extensive margin of the adoption decision (whether or not the farmer adopts) which corresponds to the empirical analysis. Initially, the farmer grows a traditional crop with a constant yield, y_a, and considers to adopt a new crop with a risky yield, $y_b = \mu + \varepsilon$, where the shock $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. As a benchmark I analyze the adoption decision under perfect information where the farmer knows the mean, μ, and variance, σ^2, of the yield of the risky crop, and without imperfections on the input market. In this setting there will be no role for the network of the farmer.

3.1 Perfect information

The farmer can choose to adopt the new crop on some share, ω, of his or her land where the total farm area is normalized to one. The total farm yield will then be a weighted average of the yield from the traditional and the new crop:

$$y = \omega y_b + (1 - \omega) y_a, \ 0 \leq \omega \leq 1$$

For simplicity I abstract from crop prices, but we could think of y_a and y_b as the value of the yields.[7] If I assume that the input cost is linear in the yield and normalize the input price to zero for now,[8] y_a and y_b represents the profits of the two crops. I return to the role of inputs in section 3.5. The crop choice of the farmer corresponds to a portfolio choice where the risk averse investor will trade-off mean and variance of the assets in his or her portfolio as exploited by Munshi (2004) in his model of acreage allocation and social learning. I follow Sargent (1979, pp.:150-151) and assume that the farmer values

[7] The analysis will be unaltered if I either assume constant prices or that the farmer only considers the current prices at the adoption decision.

[8] As long as input prices are constant, the analysis is unaffected.

the total yield according to the utility function

$$U(y) = -e^{-\lambda y}, \ \lambda > 0$$

$U(y)$ is increasing and concave and λ captures the degree of risk aversion.[9] This utility function is convenient because the expected utility can be rewritten to depend on the expected mean and variance of y, see Appendix A. The resulting expression for the expected utility is

$$E\left[U\left(y(\omega)\right)\right] = -e^{-\lambda((1-\omega)y_a + \mu\omega - \frac{1}{2}\omega^2\lambda\sigma^2)}$$

The farmer maximizes the expected utility by choosing the optimal share of farm land, ω, to allocate to the risky crop. The interior solution is found by the first order condition:

$$\omega^* = \frac{\mu - y_a}{\lambda\sigma^2} \tag{3.1}$$

This result is quite intuitive: the optimal share of land allocated to the risky crop is increasing in the difference between the mean yield and the yield of the traditional crop. It is decreasing in the variance of the crop and in the risk aversion of the farmer. For a given increase in the variance of the yield, the more risk averse farmers will choose a larger reduction in the share of land allocated to the risky crop.

Assume for practical purposes that the share of land allocated to the risky crop cannot be infinitely small. I define a share ω_{min} which is the minimum feasible value of ω other than zero. This implies that I will consider ω^* as a latent variable and the observed adoption of the risky crop will be

$$\omega = \begin{cases} 0 & if \quad \omega^* < \omega_{min} \\ \omega^* & if \quad \omega_{min} \leq \omega^* < 1 \\ 1 & if \quad 1 \leq \omega^* \end{cases}$$

Requiring a minimum share introduces the variance and the risk aversion in the extensive margin decision:

$$\omega = \begin{cases} 0 & if \quad \mu - y_a < \lambda\sigma^2\omega_{min} \\ \omega^* & if \quad \lambda\sigma^2\omega_{min} \leq \mu - y_a < \lambda\sigma^2 \\ 1 & if \quad \lambda\sigma^2 \leq \mu - y_a \end{cases} \tag{3.2}$$

[9]The Arrow-Pratt index of absolute risk aversion is $U''(y)/U'(y) = \lambda$ Pratt (1964).

3.2 Imperfect information

Now I turn to the case where the expected yield of the risky crop, μ, is unknown to the farmer. This assumption is in line with models of Munshi (2004) and Besley & Case (1993), but in contrast to the target input type models where the subject of learning is the optimal input level (Foster & Rosenzweig, 1995; Conley & Udry, 2010). For simplicity I assume that the farmers know the dispersion of the yield, σ^2, say because they are familiar with the dispersion of the rainfall. The farmer can discuss farming issues with other farmers who grow the risky crop (henceforth *informants*) to obtain information about the expected yield of the new crop. The farmer holds a belief about the expected yield:

$$\bar{\mu} \sim \mathcal{N}\left(\mu, \frac{1}{qN+k}\right)$$

I assume that the variance of the belief is inversely related to the number of informants, N, weighted by the quality of their information, q.[10] When the farmer has no informants the variance of the belief is k^{-1} which is assumed to be very large (k is a very small positive number). As the farmer discusses farming issues with more people growing the new crop, his or her belief will approach the true expected yield of the new crop.

I can find the optimal share allocated to the risky crop following the same derivations as in section 3.1, but now replacing y_b by $\bar{y}_b = \bar{\mu} + \varepsilon$. Assuming that the belief about the expected yield and the yield shock are uncorrelated, $\bar{y}_b$ will follow a normal distribution with mean μ and variance $\sigma^2 + (qN+k)^{-1}$. Thus, due to uncertainty about the expected yield the farmer will consequently overestimate the variance of the yield. The optimal (latent) share which maximizes expected utility is then equal to:

$$\omega^* = \frac{\mu - y_a}{\lambda\left(\sigma^2 + (qN+k)^{-1}\right)}$$

As expected, the optimal (latent) share of land allocated to the new crop is lower when uncertainty is introduced compared to the perfect information case (equation 3.1). The realized share will be

$$\omega = \begin{cases} 0 & if \quad \mu - y_a < \lambda\left(\sigma^2 + (qN+k)^{-1}\right)\omega_{min} \\ \omega^* & if \quad \lambda\left(\sigma^2 + (qN+k)^{-1}\right)\omega_{min} \leq \mu - y_a < \lambda\left(\sigma^2 + (qN+k)^{-1}\right) \\ 1 & if \quad \lambda\left(\sigma^2 + (qN+k)^{-1}\right) \leq \mu - y_a \end{cases} \tag{3.3}$$

Equation 3.3 suggests the first testable empirical implication:

> *Model implication 1:* Adopting the new crop is positively correlated with the number of informants growing the new crop.

When the farmer does not know anyone who grows the crop ($N = 0$) the variance of the belief about the

[10] This assumption can be motivated by a Baysian updating model where the variance of the signals from each informant is $1/q$ and the variance of the prior is $1/k$.

risky yield is very large. For sufficiently small k,[11] a risk averse farmer will not adopt a new crop which none of his or her informants grows. This is an alternative way of modeling that an information threshold has to be exceeded before adoption becomes feasible as in the model Saha et al. (1994). Discussing farming issues with just *one* farmer who grows the new crop will make the optimal latent share jump from (almost) zero to $(\mu - y_a)/(\lambda(\sigma^2 + (q+k)^{-1}))$. As long as $\mu > y_a$ the second order derivative of ω^* with respect to N is negative and hence, the extensive margin change is the unit change in the number of informants that leads to the largest increase in the propensity to adopt over the support of N.

> *Model implication 2:* The change in the propensity to adopt is larger for extensive than intensive margin changes in the number of informants.

The positive correlation between adoption and number of adopters in the network is also found in existing learning models, at least when there are few adopters in the network. Ambiguous effects for large networks are found in the target input model (Foster & Rosenzweig, 1995; Bardhan & Udry, 1999) where the subject of learning is the optimal amount of input rather than the expected yield. The farmer can learn about the optimal input both through learning by doing and learning from others which can create an incentive for strategic delay of adoption. When the farmer knows many adopters, s/he can free ride on the experimentation in the network and avoid costly experimentation on his or her own farm. This leads to an inverted U-shape relationship between the network and adoption which has been found empirically by Bandiera & Rasul (2006). In this model I do not specify how beliefs about the new crop are affected once the farmer has adopted the new crop because the empirical implications are not relevant in the context I consider.[12]

3.3 Information of different quality

Informants may not possess equally good information about the new crop. In the case of banana cultivation, RIPAT participants have received weekly training in the new cultivation technique for three years whereas non-RIPAT banana growers are likely to have less information about the new technique. When they pass on information on how to cultivate bananas I would expect information from RIPAT farmers to be of a higher quality than that of non-RIPAT farmers, $q_R > q_{nR}$. I can insert the sum of information from RIPAT and non-RIPAT informants, $qN = q_R N_R + q_{nR} N_{nR}$, into the expression for the optimal latent share of the new crop:

$$\omega^* = \frac{\mu - y_a}{\lambda\left(\sigma^2 + \frac{1}{q_R N_R + q_{nR} N_{nR} + k}\right)}$$

[11] $k < ((\mu - y_a)/(\lambda\omega_{min}) - \sigma^2)^{-1}$

[12] In the sample 97 percent of the farmers discuss farming issues with no more than three banana growers. Hence, the number of informants growing bananas is not large enough to identify a non-linear relationship between the propensity to adopt and the banana network on the intensive margin. Bandiera & Rasul (2006) find the vertex of the inverted U to be at 10 adopters in the network.

High quality informants are better at reducing the variance of the expected yield than low quality informants. Hence, they also have a larger impact on the adoption decision. This leads to the third testable implication:

> *Model implication 3*: The adoption decision is more affected by changes in the number of high quality than low quality informants.

3.4 Several risky crops

The model can be extended to include more than one risky crop. I consider the case where the farmer can choose to allocate land to two risky crops with yields y_b and y_c, which both outperform the traditional crop and hence in optimum no land is allocated to the traditional crop. The yields of the two crops are both assumed to be normally distributed and for simplicity, I assume that they are uncorrelated.

$$y_j \sim \mathcal{N}\left(\mu_j, \sigma_j^2\right), \ \mu_j > y_a, \ j \in \{b, c\}$$

Total yield is now a weighted average of the two risky crops, $y = \omega_b y_b + (1-\omega_b)y_c$. Because the two yields are uncorrelated I can simply apply the same trick as in section 3.1 and expected utility under perfect information can be written as

$$E\left[U(y)\right] = -e^{-\lambda\left(\mu_b\omega_b + \mu_c(1-\omega_b) - \frac{1}{2}\lambda\omega_b^2\sigma_b^2 - \frac{1}{2}\lambda(1-\omega_b)^2\sigma_c^2\right)} \tag{3.4}$$

First I note that the expected disutility of risk is minimized when the farmer grows both crops ($\omega_b = \sigma_c^2/(\sigma_b^2+\sigma_c^2)$) because crop diversification reduces the variance of the total yield, in particular when the two yields are uncorrelated. Next, I maximize 3.4 with respect to ω_b and the first order condition gives

$$\omega_b^* = \frac{\mu_b - \mu_c + \lambda\sigma_c^2}{\lambda\left(\sigma_b^2 + \sigma_c^2\right)} \tag{3.5}$$

Again, I let the expected yields of the two crops be unknown to the farmer and let the variance of the belief about the expected yields be the inverse of the number of informants growing the crop scaled by an information quality factor.[13] Let the number of informants growing crop b and c be N_b and N_c respectively. Assuming that the beliefs for the two crops are independently distributed I can simply insert the inflated variances in equation 3.5

$$\omega_b^* = \frac{\mu_b - \mu_c + \lambda\left(\sigma_c^2 + \frac{1}{qN_c+k}\right)}{\lambda\left(\sigma_b^2 + \frac{1}{qN_b+k} + \sigma_c^2 + \frac{1}{qN_c+k}\right)}$$

[13] For simplicity, I let the information quality factor be equal for the two crops.

and when I impose that the crops cannot be allocated a smaller non-zero share than ω_{min}, the realized share is

$$\omega_b = \begin{cases} 0 & if \quad \omega_b^* < \omega_{min} \\ \omega_b^* & if \quad \omega_{min} \leq \omega_b^* \leq 1 - \omega_{min} \\ 1 & if \quad 1 < \omega_{min} + \omega_b^* \end{cases}$$

Discussing farming issues with farmers growing crop b still increases the propensity to adopt crop b. But the question is now whether the informants who grow crop c rather than crop b affect the choice to adopt crop b? I differentiate ω_b^* with respect to N_c which yields

$$\frac{\partial \omega_b^*}{\partial N_c} = \frac{q\left[\mu_b - \mu_c - \lambda\left(\sigma_b^2 + \frac{1}{qN_b+k}\right)\right]}{\lambda(qN_c+k)^2\left(\sigma_b^2 + \frac{1}{qN_b+k} + \sigma_c^2 + \frac{1}{qN_c+k}\right)^2} \tag{3.6}$$

$$\frac{\mu_b - \mu_c}{\lambda\left(\sigma_b^2 + \frac{1}{qN_b+k}\right)} < 1 \implies \frac{\partial \omega_b^*}{\partial N_c} < 0$$

Empirically, it would appear reasonable to assume that network size is positively correlated with the adoption of new crops even after controlling for the number of people in the network who grows the new crop, simply because the network size may correlate with unobserved characteristics such as entrepreneurship and openness. However, within this model framework I present how imperfect information may lead to the opposite correlation. Network information about *another* crop makes that crop relatively more attractive through a reduction in the uncertainty about its yield. The farmer has to trade off the two crops and hence will allocate a lower share to crop b if crop c becomes more attractive. This only holds if crop b does not fully outperform crop c.[14]

I can consider crop c to represent the crop portfolio of all other risky crops grown by those of the farmer's informants who do not grow crop b. Then I can draw a final empirical implication of the model:

> *Model implication 4*: An increase in the number of informants *not* growing crop b will decrease the adoption of crop b.

3.5 Imperfect input market

Model implication 1 through 4 are derived under the assumption that the network provides knowledge about the mean yield of the risky crop(s). However, the same implications could be derived from a model with perfect information, but where the inputs are instead very costly and where the social network can lower the cost of inputs.

[14] In the case where $(\mu_b - \mu_c) / \left(\lambda\left(\sigma_b^2 + \frac{1}{qN_b+k}\right)\right) \geq 1$ then $\omega_b^* \geq 1$. Hence, crop b outperforms crop c, ($\omega_b = 1$), and ω_b is unaffected by changes in N_c.

To see how, I define the profit from growing crop b on the total farm area as $\pi_b = y_b - \kappa_b(N_b)$, where y_b represents the value of the yield, and $\kappa_b(N_b)$ is the cost of the seed input which depends negatively on the number of network members growing crop b. The yield is still risky which implies that the profit follows a normal distribution with mean $\mu_b - \kappa_b$ and variance σ_b^2. I can derive the optimal share allocated to crop b the same way as in section 3.1, so equation 3.1 is now modified by the input costs:

$$\omega^* = \frac{\mu_b - \kappa_b - y_a}{\lambda \sigma_b^2}$$

and the optimal share becomes

$$\omega = \begin{cases} 0 & if \quad \mu_b - \kappa_b - y_a < \lambda \sigma_b^2 \omega_{min} \\ \omega^* & if \quad \lambda \sigma_b^2 \omega_{min} \leq \mu_b - \kappa_b - y_a < \lambda \sigma_b^2 \\ 1 & if \quad \lambda \sigma_b^2 \leq \mu_b - \kappa_b - y_a \end{cases}$$

Consider the case where the input market is distorted by high transaction costs due to poor infrastructure, such that the cost of inputs when purchased through formal channels is so high that it is not optimal to adopt the new crop, $\mu_b - \kappa_b(0) - y_a < \lambda \sigma_b^2 \omega_{min}$. It is clear to see that allowing the network members to provide inputs at a reduced or zero cost creates a positive network effect on the propensity to adopt. This corresponds to model implication 1 above. If I furthermore assume that one network member provides a sufficient amount of inputs, then follows implication 2. Actually, there would only be an effect on adoption from an extensive margin change in the network. Alternatively, I could assume decreasing marginal returns to seed inputs for a given level of land and labor inputs which would also yield implication 2.[15] Model implication 3 would require the assumption that RIPAT network members would lower the input costs more than non-RIPAT members or be more likely to provide inputs. Given that they are obliged by the project to pass on seedlings, this is not an unreasonable assumption. Implication 4 could also easily be derived assuming that the network members growing other crops similarly lowered the adoption costs of these other crops making them more profitable than crop b.

Hence, empirical support for the four model implications would be evidence for network effects, but not for the mechanisms through which the network affects the adoption decision. It is a possibility that knowledge and input sharing simultaneously play a role.

[15] In the current specification I implicitly assume a linear relationship between seed inputs and yields, but I could instead assume a Cobb-Douglas production function which would exhibit decreasing marginal returns.

4 Data and summary statistics

4.1 Data collection

The empirical analysis is based on cross-sectional data collected in January 2011 as a part of an impact evaluation funded by the Rockwool Foundation and administered by Helene Bie Lilleør and the author (see Larsen & Lilleør, 2014). Household data were collected from RIPAT households, non-RIPAT households in RIPAT villages and households in comparison villages. This paper employs data from the choice-based sample of non-RIPAT households from the eight RIPAT villages. Households growing bananas were oversampled to ensure enough adopting households which are the households of interest for this study. Within the biomedical literature choice-based sampling is known as case-control studies and is widely used for studying infrequent events (Prentice & Pyke, 1979). The sample of non-RIPAT households consists of a random sample of households in the RIPAT villages and additional households who had received banana seedlings from RIPAT households according to RECODA records. The random sample facilitates the calculation of the population share of adopters among non-RIPAT households. The calculation is described in Appendix C. For a detailed description of the sampling scheme, see Appendix A of Larsen (2012).

The main respondent was either the person who took the decision to grow bananas or the person who takes most farming decisions, depending on whether the household had adopted bananas or not. This person was interviewed about his or her personal characteristics and network, and about the members of the household, their farm, crops, livestock, and assets. In addition, the adult female in the household was interviewed about household facilities and food security, and we collected child anthropometrics.

The sample of non-RIPAT households for the analysis of the adoption decision is constructed as follows: 597 non-RIPAT households in the eight RIPAT villages were interviewed in total. Out of these, 62 households are disregarded due to missing data on network or other explanatory variables. The data are not systematically missing from either adopting or non-adopting households. In addition, 26 households are left out of the analysis because they either claim to have planted their first improved bananas before 2006 or because they moved to the village later than 2006 when RIPAT was implemented.[16] This leaves a final sample of 509 households among which 193 are growing improved bananas as listed in the final row of table 1.

4.2 Measuring networks

When assessing the role of the network for adoption of technology, it is important how the network is measured. Maertens & Barrett (2013) argue that the network is almost surely misrepresented if the researcher does not have explicit network data, but instead relies on proxy measures such as other farmers

[16] Including households who planted before 2006 as either adopting or non-adopting households does not alter the results. Neither does the inclusion of immigrants.

in the village (e.g. Foster & Rosenzweig, 1995; Munshi, 2004; Moser & Barrett, 2006), or geographical neighbors of the farmer (e.g. Krishnan & Patnam, 2014). Direct network measures typically capture egocentric networks either by prompting the farmer about links to other farmers in the study (Conley & Udry, 2010; Carter et al., 2014; Magnan et al., 2015) or by open-ended questions about whom they discuss farming issues with (Bandiera & Rasul, 2006). Open-ended questions might only elicit the farmer's "strong" network links because the "weak" links may be forgotten when the farmer is not prompted (Maertens & Barrett, 2013). However, the network size is not captured when the researcher only asks about links to other farmers in the sample, and I show in this paper that the size is important to account for. Also, women may have systematically smaller network measures than men if only the network links within the village are elicited and it is a patri-local society where the women moves to the men's villages when they marry. This is the case in the context I study.

Hence, I employ three open-ended questions to capture the egocentric networks of non-RIPAT respondents. The phrasing below was used for non-adopting farmers while adopting farmers received the questions in past tense and the recall frame in square brackets was used.

Network size: Think about your relatives and friends and other people that you know. [Before you decided to start growing improved bananas,] how many people do you discuss farming issues with?

Banana network: Among these, how many of them are growing improved bananas [before you decided to grow improved bananas]?

RIPAT banana network: If any of these are RIPAT farmers, could you please give me their names?

The timing of the recall was chosen to avert the potential upward bias due to endogenous network formation: If the network was measured after the adoption decision I could capture links between banana cultivating farmers that were established because they both grow bananas. This would induce an upward bias in the correlation between the banana network and the adoption decision.

The questions are sequential such that the mentioned farmers will be a subset of the response to the preceding question. This implies that only RIPAT farmers who were growing bananas were listed. The listed RIPAT farmers can be linked to detailed household and farmer characteristics, because the data collection also covered all RIPAT farmers. I exploit this information in section 6.1 to study whether the effect of the RIPAT network depend on the socioeconomic characteristics of the RIPAT farmers.

The empirical network measures easily relate to the network in the theoretical model presented in section 3. The banana network net of the RIPAT banana network will capture the number of non-RIPAT banana growers in the network. From model implication 1, I expect that the RIPAT and non-RIPAT network are positively correlated with the propensity to adopt. Model implication 3 suggests that the RIPAT network has a stronger impact than the non-RIPAT network. The network size for a given banana network will capture the number of informants growing other crops than bananas and model implication 4 predicts a negative correlation between the network size and the propensity to adopt conditional on the banana network. Furthermore, controlling for the network size ensures that the impact of the banana network is not confounded by the network degree.

Is there any correlation between the banana network measured and the propensity to adopt in the raw data? Figure 2 shows the sample share of adopting households depending on the number of people in the farmer's banana network (2a) and the farmer's RIPAT network (2b).[17] The sample share of adopting farmers is clearly larger for the subsets of farmers who discuss farming issues with banana growers which corresponds with model implication 1. The figure suggests that the greatest difference is on the extensive margin, i.e. whether you discuss farming issues with at least one RIPAT farmer or other banana grower. This is in line with model implication 2.

4.3 Summary statistics

Table 1 summarizes farmer and household characteristics for the full sample of non-RIPAT households and for adopting and non-adopting households separately. Due to the choice-based sampling of non-RIPAT households the adoption share in the sample is 38 percent which is almost twice the population share of 20 percent of adopting households among non-RIPAT households.

The explanatory variables of interest are the network variables. The farmers in the sample discuss farming issues with 2.8 people on average, where 0.5 are RIPAT banana growers, 0.3 are non-RIPAT banana growers and the remaining two people do not grow bananas. Adopting farmers are more likely to discuss farming issues with banana growers than non-adopting farmers, and though adopters also have a larger total network size than non-adopters, they discuss farming issues with 0.8 fewer people who are not growing bananas.

Furthermore, I control for a range of farmer and household level characteristics. At the farmer level, I include gender, age, religion and literacy. The reference category for the religion dummies is that the farmer is Protestant. 'Other religion' is a combined group of both traditional religion practitioners, Seventh Day Adventists and other groups that do not fall into the three main religion groups. Adopting and non-adopting farmers are quite similar, though adopting farmers are slightly more likely to be Catholic and less likely to be Muslim than non-adopting farmers.

At the household level I consider different components of the household structure, namely the highest education level obtained, available household labor, whether the household head is a widow(er), the wealth of the household, and the farm size. This range of variables address constraints to adoption with respect to inputs to agricultural production: capital, labor, human capital, and land. The highest level of education achieved within the household will be the level of formal knowledge that the farmer can tap

[17] Recall that the adoption share in the sample does not correspond to the population share due to choice-based sampling.

into. Since the highest level is 'completed primary education' (7 years) in 58 percent of the households I use this as the reference point and include indicators for having less or more education than completed primary. Adopting households are more likely than non-adopting households to have a household member with more than primary education. To capture household composition I control for whether the household head is a widow(er) and for the available household labor which is measured as the number of household members who can do hard manual labor to full extent. Adopting households have significantly more household labor and are less likely to be widowed. The level of wealth of the household is measured by a Tanzanian poverty score developed by Schreiner (2011) and I also include the number of acres of land the household employ in 2006. I use a recall measure for the farm size since it may be endogenous to the adoption decision, say if a farmer finds that banana cultivation is lucrative and rent in more land.[18] I do not have a recall measure for the poverty score but in the impact evaluation of RIPAT (Larsen & Lilleør, 2014) we do not find the poverty score to be significantly affected by project participation. Adopting households are significantly more wealthy than non-adopting households, but they do not have more land. In addition I measure remoteness by the distance from the GPS location of the household to the nearest road,[19] and this measure is not significantly correlated with adoption.

I further include variables that capture agricultural practices, entrepreneurship and growing conditions that may correlate with both network and adoption.[20] Entrepreneurial households who are open for change could be more likely to participation in an NGO project (other than RIPAT) and to grow more different crops (net of improved bananas), and these two variables are indeed positively correlated with adoption. Whether or not the household has grown traditional bananas indicates if the household has some prior knowledge about or special preferences for banana cultivation and it appears to be an significant determinant of adoption: Adopting households are 18 percentage points more likely to have grown traditional bananas than non-adopters. This is important to control for as farmers who have grown traditional bananas may be more likely to discuss farming issues with each other. The local growing conditions are captured by the number of banana growers within a radius of 0.5 kilometers from the household where the distance is measured as the distance between GPS points taken at the household's compound. As we have not collected census data the measure is not complete, but it is a good proxy for the growing conditions that the household faces.[21] Indeed, there is geographical clustering in the adoption of banana cultivation with adopting households having more neighbors who also grow bananas than non-adopters. The inclusion of this variable may cause the network estimates to be downward bi-

[18] As 97.5 percent of the sample owns at least some of their land and 83.9 percent owns all of the crop land they cultivate, inadequate incentives with respect to farm tenure arrangements should not be a constraint. Hence, I do not distinguish between whether the household owns or rent in the land that they cultivate.

[19] Data on roads are downloaded from OpenStreetMap (http://download.geofabrik.de/africa/tanzania.html) and kilometer distance from household GPS points is calculated in ArcGIS.

[20] I could further include measures to capture access to information such as household ownership of a mobile phone or a radio, but these variables are uncorrelated with adoption and inclusion of them does not alter the results. I leave them out to reduce dimensionality.

[21] Once the number of banana growers within a 0.5km radius is controlled for, use of irrigation channel, historical rainfall at the household level and distance to nearest waterway becomes insignificant. Hence these measures are not included. Including them does not alter the results.

ased if farmers mainly discuss farming issues with their neighbors. However, if I exclude it, the network estimates may be confounded by a correlation in growing conditions within networks, hence I prefer the conservative estimates.

The eight villages in the data have a total of 24 subvillages with two to five subvillages in each village. Between six and 55 households are included in the sample from each subvillage. One subvillage has no adopting farmers and hence, these six observations are excluded when subvillage fixed effects are controlled for.

5 Econometric analysis

I estimate a logistic model of the probability of adopting banana cultivation:

$$Pr\left\{adopt_i = 1\right\} = \Lambda\left[\beta_1 R_i + \beta_2 nR_i + \beta_3 N_i + \delta X_i + \gamma G_i + \alpha_s\right] \tag{5.1}$$

where R_i is the RIPAT banana network of farmer i, nR_i is the non-RIPAT banana network and N_i is the network size. According to model implication 1 from section 3 β_1 and β_2 are positive, while model implication 3 predicts that $\beta_1 > \beta_2$. Model implication 4 suggests that β_3 is negative. In the first specification the two banana network variables are specified as the number of (non-)RIPAT banana farmers in the network. In a second specification I let R_i and nR_i be vectors of three indicator variables: Discuss farming issues with at least one; at least two; or at least three (non-)RIPAT banana farmers. This specification will allow me to test the model implication 2 that the extensive margin network effect is larger than the intensive margin effect.

In addition to the network variables, I control for farmer and household characteristics, X_i, and growing conditions, G_i, as these may both correlate with the network and the adoption decision. These are described in section 4.3. I can further control for local factors that makes adoption behavior correlate within the subvillages by subvillage fixed effects, α_s.[22] All standard errors are clustered at the subvillage level.

The logistic model is based on the assumption that the individual unobserved characteristics are logistically distributed. This is a convenient model for choice-based samples because it provides consistent estimates of the parameters–apart from the constant term–as opposed to the linear probability model and the probit model (McFadden, 1973; Prentice & Pyke, 1979). The bias of the constant term can be corrected if the proportion of adopting households in the population is known. This enables the calculation of marginal effects when subvillage fixed effects are not included. Appendix B derives the consistency of the logit estimator for a choice-based sample and subvillage fixed effects. It further describes how the correction of the constant term is calculated.

[22]The model including fixed effects are estimated using conditional maximum likelihood where I only use within-subvillage variation in the adoption behavior to estimate the parameters (Chamberlain, 1982).

Marginal effects are calculated for the specification that includes farmer and household characteristics but not subvillage fixed effects. For count variables such as the network variables, the marginal effect is calculated as the change in the propensity to adopt for a discrete change around the mean value of the count variable.[23] The remaining explanatory variables are evaluated at the sample mean and the constant is corrected for choice-based sampling. In the second specification I calculate For indicator variables I provide the marginal effect of a discrete change and for continuous variables I provide the usual marginal effect, still correcting the constant term and using the sample mean of the remaining variables.

The model in section 3 motivates a causal interpretation of the β estimates as network effects. However, the identification of such network effects requires careful scrutiny of all potentially confounding effects and considerations of reverse causality. The estimates of the network effects may be confounded by contextual or correlated effects using the terminology of Manski (1993), and they may be driven by self-selection into RIPAT. I will discuss the issues of identification in section 6 and address them with three falsification tests. I will also discuss whether the network effects are driven by dissemination of information or provision of inputs. But before I address the issues of identification I will present the regression results based on the specification in equation 5.1.

5.1 Empirical results

Table 2 presents logistic coefficients and marginal effects for the propensity to adopt improved banana cultivation. Column (1) presents the simple logistic regression of the propensity to adopt on the three network variables. Discussing farming issues with a banana grower – whether RIPAT or non-RIPAT – is significantly positively correlated with the decision to adopt which is in line with model implication 1. Knowing an extra RIPAT farmer appears to be five to six times as important as knowing an extra non-RIPAT farmer who grows bananas, given network size, which correspond to model implication 3. The network size is negatively correlated with adoption when the number of banana growers is controlled for as predicted by model implication 4. These strong correlations persist when I include farmer and household characteristics in the regression. Furthermore, they remain unaffected when I account for subvillage fixed effects.[24] Since these parameters are only identified by variation within subvillages, factors that cause adoption rates to be correlated within subvillages such as soil quality, distance to markets and village institutions are not confounding the network effects.

The marginal effect show that the RIPAT network is really economically significant: Knowing an extra RIPAT banana grower increases the propensity to adopt by 24 percentage points. The non-RIPAT

[23] For the number of RIPAT and non-RIPAT banana growers in the network this corresponds to a change from zero to one while for the network size it is a change from two to three. For the second specification, I consider discrete changes in the indicator variables. E.g. the marginal effect of knowing at least two RIPAT banana growers is calculated by changing this variable from zero to one while *knowing at least one RIPAT banana grower* is set equal to one and *knowing at least three* is equal to zero.

[24] The number of observations is reduced by six households because one subvillage does not have any adopting farmers in the sample.

network appears to provide information of much lower quality as an extra farmer only increases the propensity to adopt by 5 percentage points. On the other hand discussing farming issues with a person not growing bananas reduces the propensity to adopt by 5 percentage points. These results illustrate well a situation of information deficit: Farmers are easily convinced to try a new crop by well-informed farmers, they are less affected by farmers who provide second hand knowledge and if they are in general more informed through a larger network, they are more difficult to persuade. But they might as well be explained by the provision of seedlings through the network which I return to in section 5.2.

Turning to farmer characteristics, female farmers are 14 percentage points more prone to adopt banana cultivation than male farmers. This is well in line with anthropological field work in the area which concludes that women generally have the authority over bananas as compared to beans which is the domain of men (Mogensen & Pedersen, 2013). There appears to be an inverse U-shaped relationship between the propensity to adopt and the age of the farmer, though the two terms are not jointly significant at the ten percent level.[25] Though the coefficients are not significant, the pattern is well interpretable. Until the age of 41 there is an increasing relationship between the farmer's age and adoption while the relationship is negative for older farmers. This can be explained by the different phases in a household where a young farmer has to spend time on child rearing, while when the children become older the household can draw on teenage labor force. For older farmers, the children may have left home leaving fewer hands in the family farming activity.

Religion appears to play an important role showing that Catholics are 19 percentage points more likely to adopt than Protestants who constitute the reference group.[26] The other religion dummies are not significantly different from zero. The literacy of the farmer does not correlate with the adoption decision.

The highest education attained in the household does also not correlate with the adoption decision.[27] The little importance of education suggests that the new technology is so simple that lack of formal education is not a barrier to adoption. On the other hand, household labor appears to have some impact on adoption though it is only significant at the ten percent level. It is measured as the number of household members who are able to do hard manual labor to a full extent. As the establishment of a banana plantation requires a lot of hard manual labor it is intuitive that the available household labor is positively correlated with adoption. Whether the household head is a widow(er) seems to be negatively correlated with the adoption decision as expected. The estimated coefficient is rather large, but quite inaccurate and hence not statistically significant. Naturally, a widow(er) household has less available household labor, and the strong negative correlation between household labor and the widow

[25] Wald tests of joint significance: $\chi^2_{(2)} = .49$, $p = 0.11$.

[26] Catholics are equally represented among RIPAT and non-RIPAT farmer, so the large coefficient can not be explained by Catholics being reluctant to join RIPAT groups. The role of religion in networks would be an interesting topic for future studies.

[27] Neither does the education of the farmer if that was included instead.

dummy explains the large standard errors. The wealth of the household as measured by a poverty score does not appear to be a determinant of adoption and neither does the number of acres the household has access to. Hence, little wealth or limited access to land does not seem to be important barriers to adoption, supporting the trialability of banana cultivation and further suggesting that network effects are not driven by access to credit. The distance to the road does not correlate significantly with adoption so more remote farmers are not more or less likely to adopt banana cultivation. Bandiera & Rasul (2006) finds participation in other NGO projects to be an important determinant of adoption of sunflower cultivation, and in Larsen & Lilleør (2014) we find project participation in the past to be correlated with participation in RIPAT. However, I do not find project participation to correlate with adoption of banana cultivation among the non-RIPAT participants.

The last three variables cover agricultural practices and conditions. I include the number of crops the household grew in 2010, net of traditional and improved bananas, to control for the combination of entrepreneurship and preference for risk diversification that would induce the farmer to plant many different crops. The number of crops grown in 2010 is indeed positively correlated with the adoption of banana cultivation. Previous or current cultivation of traditional bananas indicates that the household has some prior knowledge about banana cultivation reducing the information gap. It could also capture that the household has adequate growing conditions or special preferences for banana cultivation. Households who grows or have grown traditional bananas are 17 percentage points more likely to adopt improved banana cultivation. Finally, I control for growing conditions such as soil quality and rainfall by including the number of banana growers within a radius of 0.5 kilometers and the parameter is positive and significantly different from zero though less so when subvillage fixed effects are included as they capture some of the same variation in the data.

Among the list of characteristics, the number of RIPAT banana growers in the network of the farmer prevails as one of the most important determinants of adoption both economically and statistically. It represents the highest marginal effect on the propensity to adopt and the t-statistic of the parameter estimate of 6.00 is by far the largest t-statistic of the included controls. As the sample share of adopting households is 24 percent for those farmers who do not know any RIPAT farmers, discussing farming issues with just one RIPAT farmer doubles the propensity to adopt.

I use standard errors clustered at the subvillage level to assess the significance level of the estimates. However, there are 24 subvillages in the data and this rather low number of clusters raises the question of the asymptotic distribution of the test statistics. Following Cameron et al. (2008) I address this question by estimating a linear probability model using ordinary least squares and calculate wild bootstrap-t p-values for the network variables. The coefficients to the RIPAT network and the network size both have p-values below 0.01, however the non-RIPAT network variable is not statistically significant when all covariates are included. Results are shown in panel A of Table A.1 in Appendix.

Even though I address the oversampling of adopting households by using the logit estimator, there might be an additional concern. The logit estimator provides consistent estimates if the additional adopting households are drawn randomly among all adopting households in the population. I identify additional adopting households through RECODA records and may thereby exactly sample households who are connected to RIPAT farmers potentially leading to an upward bias in the network estimates. However, I obtain the same parameters when I only include households drawn randomly, see Table A.2 in Appendix. If anything, the network estimates based on the random sample are larger, so I do not overestimate the effects using the non-random sample.

Figure 2 suggests that the relationship between the propensity adopt and the banana network is not linear, but rather that the extensive margin change in the number of banana informants is what matters for adoption. This is also supported by model implication 2. The regression results presented in Table 3 allows for a flexible relationship between the propensity to adopt and the two banana network variables. Each banana network variable is split into three indicator variables: Discuss farming issues with at least one; at least two; or at least three (non-)RIPAT banana farmers.

The correlation between the banana network and adoption is clearly driven by the extensive margin: Knowing at least one RIPAT banana grower increases the propensity to adopt by 37 percentage points, corresponding to a 154 percent increase in adoption! When controlling for having at least one (non-)RIPAT banana grower in the network, the second and third banana grower does not correlate with the adoption decision. Only when subvillage fixed effects are included the second RIPAT farmer has a marginally significant effect on adoption and the size is less than two thirds of the extensive margin effect.[28]

If I believe that the network effects are indeed driven by dissemination of knowledge, this result tells me something about the nature of the information constraint that the farmers are facing. It suggests that the informational barrier is relatively easily surpassed as only *one* source of information is needed to relax the constraint. This requires that the agricultural technique for improved banana cultivation is relatively simple to learn, and that just one observation can convince the farmer that improved banana cultivation is very profitable compared to many of the traditional crops grown in the area.

However, the importance of the extensive margin of the network is also very well in line with the idea of input provision through the network. It only requires one RIPAT contact to get hold of the first improved banana seedlings which makes adoption of improved banana cultivation feasible. This would generate the network effects found even if the farmers do not face an information constraint about the agricultural technique or profitability of improved banana cultivation. Both the input and information channel could very well be in play at the same time, and I am not able to fully disentangle the two. In

[28] According to wild bootstrap-t p-values from OLS regressions the extensive margin of the RIPAT network and the network size are both significantly different from zero at the five percent level at least, while the non-RIPAT banana growing network is insignificant (see panel B of Table A.1 in Appendix)

the following section I present evidence that provision of inputs takes place in the networks.

5.2 Provision of inputs

There could be several interpretations consistent with the network effects found. The theoretical model presented in section 3 demonstrates that the network effects found are consistent with a story of social learning. However, as pointed out in section 3.5, the same model implications could be derived from a model where the network members provide free access to inputs, but no information. Hence, the fact that the data are in accordance with the model from section 3 does not allow me to conclude on whether the network effects are driven by learning about the expected yield or reduced costs of adoption.

Data on where adopting farmers got hold of their first improved seedlings can shed a little light on the role of input provision. Adopting farmers were asked who gave or sold them the first banana seedlings and if they mentioned a RIPAT farmer I can cross check if the farmer is also in the network. The data are presented in Table 4. The first row covers the full sample of adopting farmers and shows from whom they received their first banana seedling. The majority (37 percent) received it from a RIPAT farmer in their network, while 29 percent received it from a RIPAT farmer who is not in their network.[29] Hence, two thirds received the seedling from a RIPAT farmer. A quarter received it from a non-RIPAT farmer and the remaining nine percent either received it directly from the implementing organization, a RIPAT group, or another NGO. No one bought the seedlings through formal channels suggesting that limited access to improved banana seedlings could be a binding constraint for adoption.

Network connections could also affect the price of the banana seedlings, and I can use information on whether the farmer paid in cash or in kind for the first seedling(s) to shed light on the issue. The majority (78 percent) received the first seedlings for free and this is statistically independent from whether they received the seedling from a RIPAT or non-RIPAT farmer. On the other hand, 44 percent of those who received their seedlings from the other sources mentioned above had to pay either in cash or in kind for the seedlings. This suggests that being connected to other banana growers reduces the cost of inputs.

The networks are measured by asking whom the farmer discuss farming issues with which relates to the dissemination of information. But are these network measures correlated with the provider of the first seedling? Not surprisingly, discussing farming issues with one or more RIPAT banana growers increases the probability of receiving a seedling from a RIPAT farmer. This can be seen from the second row of Table 4 where I only consider adopting farmers who know at least one RIPAT banana grower (107 observations). Among these, 65 percent receive the first seedlings from a RIPAT network member suggesting that the network plays an important role for the provision of inputs. Farmers who discuss farming issues with at least one non-RIPAT banana grower (third row) are also more likely to receive the first seedling from a non-RIPAT banana grower compared to the average adopting farmer (first row).

[29]The person may belong to the higher order network.

Hence, the provision of seedlings through networks could be a plausible explanation for why the network effects are found to be so large. In particular, the fact that the network effects are dominant on the extensive margin could be driven by seedling provision. You may only need to know *one* banana grower to get your first improved seedlings.

The difference in the network effect for RIPAT and non-RIPAT banana growers could also be explained by seedling provision. Among the adopting farmers who knows both at least one RIPAT and non-RIPAT banana grower, 65 percent receives the first seedling from a RIPAT farmer (either in network or not) while only 29 percent receives the seedling from a non-RIPAT farmer. This suggests that RIPAT farmers are more likely to share seedlings in their network. There could be several explanations for this. It takes approximately one year from the establishment of a banana plant before the farmer can harvest seedlings that can be passed on to other farmers in the network. Hence, if non-RIPAT banana growers have planted recently, they may not be able to share seedlings in their network. Also, RIPAT farmers may be more prone to share seedlings than non-RIPAT farmers. In fact, 61 percent of the RIPAT farmers who grow bananas and have passed on seedlings to other farmers mention "obligation in the project" as one of the reasons for passing on improved banana seedlings to other farmers. This points to the importance of the solidarity chain principle for the diffusion of improved banana seedlings.

This fact raises the question of whether the adopting farmers simply plant a few banana plants because they received the seedlings as gifts which leads to the high impact of RIPAT network on adoption or whether they really learn the new technology and adopt it because they perceive improved banana cultivation to be advantageous. I investigate this issue by considering whether the farmer has a plantation with more than ten banana plants instead of at least one banana plant. I find the same network effects on the propensity to establish a banana plantation as on propensity to grow at least one plant (see Table A.3 in Appendix). Hence, the network effects on adoption are not artificially high because the solidarity chain principle could induce RIPAT farmers to pass on banana seedlings to other farmers who where not interested in banana cultivation.

Naturally, data on seedling provision can only explain variation within adopting farmers as they do not exist for non-adopting farmers. To assess the constraints faced by the non-adopting farmers, they were asked why they had not planted improved bananas, and Table 5 presents the categories of answers to this open-ended question. Water shortage is the dominant self-reported reason for not planting improved bananas, while land constraint is the second most important reason mentioned by the farmers. Lack of knowledge about production techniques is mentioned as frequently as no access to seedlings suggesting that these two constraints are both in play and are equally important.

The data do not allow me to draw a firm conclusion on the drivers of the network effects found, but there is suggestive evidence that the network provides a necessary input for improved banana cultivation, namely the improved banana seedlings. The provision of inputs through the network does not rule out that dissemination of knowledge is also taking place. The fact that I find very strong network effects could suggest that both channels are at work.

6 Identification of network effects

Identification of social interaction effects is inherently difficult because most networks are endogenously shaped through individual choices. This may cause behavior to correlate within the network for other reasons than social interaction In section 5.1 I found that the network of the farmer is a very strong predictor of the farmer's adoption decision. Regardless of whether this network effect is driven by information or input provision I consider it to be a social interaction effect as the adoption behavior of the network member is a prerequisite for both the information and input channel. In this section I will go through all the different causes of correlated adoption behavior in the networks which cannot be assigned to social interaction and address them one by one.

In order to identify social interaction, Manski (1993) employs the useful vocabulary of endogenous social interaction effects, contextual effects and correlated effects. The *endogenous effects* describe how the behavior of the individual is affected by the behavior in the peer group. These are the network effects I want to identify. The *contextual effects* (or exogenous social effects) cover how the behavior of the individual is affected by the exogenous characteristics of the group such as education or wealth.[30] I investigate whether the network effects found are driven by characteristics of the network rather than their adoption behavior in section 6.1. The *correlated effects* are covariation in behavior within a group due to similar unobserved individual characteristics or because group members face a similar environment. Endogenous network formation naturally leads to a correlation in individual unobserved characteristics if similar people prefer to share information. In addition, behavior may be spatially correlated due to growing conditions or institutions. I address these confounding factors in section 6.2.

Furthermore, Manski (1993) discuss the *reflection problem* that arises when the researcher wants to determine how the average behavior in a group affects the individual behavior in the group. The simultaneity within the group makes it difficult to identify who affects whom. In the network I study there is a natural ordering of events which circumvents this kind of simultaneity bias. I consider how non-RIPAT farmers are affected by discussing farming issues with RIPAT farmers who have adopted improved bananas. The ordering is created by the fact that RIPAT farmers were the first to be introduced to improved banana cultivation. Data on time of adoption allows me to check that non-RIPAT farmers did indeed plant bananas later than the RIPAT farmers in their networks. In 96 percent of the links, the RIPAT farmer planted before the non-RIPAT farmer and for 98 percent of the adopting non-RIPAT farmers at least one of the RIPAT farmers in their network planted before them. Hence, I do not consider

[30] Manski (1993) uses the word "exogenous", however, it should be noted that if networks are endogenously formed then the characteristics of network members may be endogenous too.

simultaneity bias to be a great concern.

Another type of reverse causality could occur if the network was measured after adoption as banana growing farmers may endogenously form networks after adoption. However, recall that I capture the network *prior* to the adoption decision. Though I cannot rule out the existence of a recall bias, this suggests that the estimated network effects presented in section 5.1 are not confounded by the creation of links between banana cultivating farmers after they have chosen to adopt.

Even though networks are measured prior to the adoption decision, the self-selection into RIPAT could potentially cause adoption behavior to correlate without the RIPAT farmer affecting the non-RIPAT farmer to adopt. I discuss this issue in section 6.3.

It should be noted that when studying the impact of the network on adoption behavior, I cannot distinguish imitative behavior from learning as pointed out by Foster & Rosenzweig (1995). It would require farm productivity data to distinguish between imitation and learning which are very costly to collect and often subject to a large degree of measurement error. Nevertheless, if I assume that farmers are rational, adoption must indicate that they perceive improved banana cultivation to be relatively advantageous either with respect to profits, household food security or social factors such as prestige.

6.1 Contextual effects

Could the network effects found in section 5.1 be driven by the characteristics of network members? For instance, if banana growers are on average wealthier than other farmers, then knowing several banana growers implies knowing several wealthy people who may provide informal credit or insurance for you. In that case, a positive correlation between the number of banana growers in network and own adoption is not an evidence for learning or input provision but confounded by access to informal credit.

Ideally, I would like to control for the average characteristics of all the information network members to ensure that the correlation between adoption and the adoption behavior in the network is not driven by exogenous characteristics of network members. However, I only have detailed information about the RIPAT farmers in the network and not other network members.[31] To the extent that exogenous characteristics are highly correlated within the network, controlling for farmer characteristics, X_i, that are expected to affect adoption partly resolves the issue. But it is not sufficient in the case of heterogeneous networks. Though I do not have data on all network members, I can exploit the detailed data I have on RIPAT farmers. I split the RIPAT farmers in the network based on five central socioeconomic characteristics: wealth, land, education, gender, and age, and I examine whether the network effects differ dependent on the characteristics of the network members. If the network effects were driven by the characteristics of the farmers in the network rather than their adoption behavior (contextual effects), I would expect to find different network effects for e.g. rich and poor network members. No differential

[31] To my knowledge, the only paper analyzing adoption of agricultural technologies and networks with detailed information on all network members is that of Van den Broeck & Dercon (2011).

effects would support the hypothesis that the network effects found are not driven by contextual effects.

I measure wealth by a poverty score with a range of 0-100 (Schreiner, 2011) and split the sample of RIPAT farmers in networks at the mean poverty score, 47.4. I split the RIPAT farms into small and large by the average number of acres of 4.4. With respect to education, RIPAT farmers are divided into three groups: less than seven years of education (26.9 percent), seven years of education (68.1 percent) and more than seven years of education (5.0 percent of the sample). The gender split is self-explanatory, while the sample is split into young and old farmers at the mean age of 46.9 years.

Table 6 shows the estimation results together with tests of equal network effects across different characteristics of network members. Column (1) does not provide support for the hypothesis that the network effect is driven by access to credit through network members as the estimated effect of knowing rich RIPAT banana growers is in fact lower than knowing poor RIPAT banana growers. However, this difference is very far from being significant. The same pattern shows when I split the RIPAT network on the size of the farm in column (2) which could also be considered as a measure of wealth. RIPAT farmers with a small farm actually appear to have a stronger network impact, but again the difference is not significant.

Turning to the split on farmers' education in column (3), it appears that there is a smaller effect from knowing RIPAT banana growers with less than seven years of education, but there is no significant difference on the three estimated network effects for different education categories. The coefficient for knowing RIPAT farmers with a high education is imprecisely estimated as the group is fairly small.[32] To ensure that the acceptance of the null hypothesis is not driven by large standard errors induced by the high education category, I combine the high and medium education category, and again I accept the null of no difference in network effects across education of network members.[33]

As can be seen in column (4), the impact of knowing a male RIPAT farmer seems to be larger than knowing a female RIPAT farmer. Nevertheless I again reject that the network effects are differential across gender. Neither do the estimates and test results in column (5) provide evidence for a difference in the network effects across age.

Hence, I conclude that the network effects appear to be rather homogenous across these five socioeconomic characteristics which indicates that the network effects are not driven by the characteristics of the farmers in the network, i.e. contextual effects. At least, the network effects do not seem to be driven by access to informal credit or e.g. by knowing older RIPAT farmers who are maybe more respected and influential in the village.

[32] Only 13 farmers in the sample know a RIPAT farmer with high education.

[33] Estimation results not reported. Wald test of equal effects for knowing RIPAT farmers with low and medium/high education: $\chi^2_{(1)} = 1.00$, $p = 0.318$.

6.2 Correlated effects

I distinguish between correlated effects due to environment (growing conditions and location institutions) or individual unobserved characteristics.

Farmers within a network may behave similarly because they face the same environment. Agricultural activities may be correlated for neighboring farmers due to similar growing conditions rather than social network effects. If the subvillage leadership is supporting and promoting banana cultivation in a particular subvillage then a correlation in adoption behavior within networks in the subvillage would not necessarily indicate the existence of social network effects. In the empirical specification in section 5.1 I have addressed these issues in several ways.

I capture the growing conditions of a farmer by the number of banana growers in my sample within a radius of 0.5 kilometers from the farmer's dwelling as measured by GPS,[34] and the results in Table 2 show that this measure is an important determinant of adoption. As all RIPAT farmers are interviewed and in some villages all identified adopting farmers are interviewed, this measure almost corresponds to the actual number of banana growers within a radius of half a kilometer. However, in the villages where adoption is very wide spread so that the sample does not include all adopting households in the village, it understates the number of adopters within the radius. This is somewhat problematic since it will not capture the full effect of growing conditions in these villages. To mitigate this problem, I could additionally control for the historical rainfall within one square kilometer of the household,[35] the distance from the household to the nearest waterway[36] and whether the household uses an irrigation channel. However, I do not find any of these measures to be important for adoption once the number of neighboring adopters is controlled for, neither does inclusion of them affect the estimated network effects. Hence, I consider the number of adopters within a small radius to be a good measure for the growing conditions of the farmer. To ensure that institutional effects are not driving the results I show that the network estimates are invariant to the inclusion of subvillage fixed effects. The fixed effects also capture general equilibrium effects such as the effect of wide spread adoption in the local market price of bananas.[37]

Another important correlated effect stems from the likely correlation of unobserved individual characteristics within networks which are formed by individual choices. Entrepreneurial farmers may first of all have larger networks and hence, be more likely to know adopting farmers. Thus, I control for network size in all regressions. In addition, an entrepreneurial farmer may choose to discuss farming issues with

[34]The distance is calculated using the 'geodist' package in Stata. The GPS measure is taken at the household dwelling and not at the farmer's plot(s), but this should not add too much noise as the majority of households have plots that are contiguous to their dwelling.

[35]For historical rainfall I use interpolated data on yearly precipitation measured in mm from the period 1950-2000 available from http://www.worldclim.org/ and link it to the households using GPS coordinates.

[36]Data on waterways is downloaded from OpenStreetMap available from http://download.geofabrik.de/osm/africa/ and the kilometer distance to household GPS points is calculated using ArcGIS.

[37]However, it should be noted that the majority of farmers face periods of food insecurity and mainly grow bananas for home consumption.

other farmers who are themselves entrepreneurial. Hence, a correlation between their adoption behavior may simply reflect that they are of the same type rather than being an indication of social interaction effects. If eligibility into RIPAT had been randomized, I could have used the random variation in the network of the non-eligible farmers to circumvent this problem (see e.g. Kremer & Miguel, 2007). But because participation in RIPAT was voluntary I must address the potential correlation of unobservables within the network. I do that by performing the following placebo study.

6.2.1 Placebo study

To examine if the strong correlation I find between adoption behavior and adopters in network is driven by a correlation in unobservables I consider adoption of three other crops: vegetables, sunflowers and sugarcanes which are all profitable cash crops.

Cultivation of vegetables (e.g. onions, tomatoes) is very profitable but also requires access to water and intensive seasonal labor input. Sunflowers can be grown under rather dry conditions and the sunflower oil can be extracted from the seeds with a simple hand press. Sugar cane is a perennial grass that can be grown under varying conditions but access to irrigation water increases yields.[38] If the profitability of banana cultivation dominates the profitability of vegetables, sunflowers and sugarcanes for all farmers then the placebo test has no bite. However, I would argue that this is not the case. Farmers who have access to plenty of water would profit more from vegetables than bananas whereas it might be more beneficial to grow sunflowers for farmers who have very limited access to water. I have chosen these three crops because their profitability relative to banana cultivation varies across farmers conditional on their available inputs. Banana cultivation is not very likely to dominate the cultivation of all of these crops.

If the correlation between the number of banana growers in network and adoption of banana cultivation is driven by a correlation of entrepreneurship in the network I would expect the number of banana growers in network to explain variation in the adoption of vegetables, sunflowers and sugarcanes. If knowing more RIPAT farmers is simply a proxy for being more open to new ideas then it should be correlated with the adoption of other crops too. However, if the RIPAT banana growers in the network also grew vegetables a positive correlation would occur in the case of social interaction within vegetables production. Hence, I control for the number of RIPAT banana growers in network who *also* grow the placebo crop. Since only 13 percent of the farmers grow sunflowers and eight percent grow sugarcane there are several subvillages with no variation in the adoption of the placebo crop which results in a fewer number of observations.

Table 7 presents the estimates from logistic regressions of adoption of vegetables, sunflowers and sugarcanes, respectively, on the network variables and farmer and household characteristics. None of the network estimates are significantly different from zero and they appear small in magnitude. These results show that the number of banana growers in the network cannot explain adoption of any of the three

[38] Information on cultivation of vegetables, sunflowers and sugarcanes is based on conversations with Jens Vesterager, Programme Manager, Rockwool Foundation.

placebo crops.

The lower number of observations in particular in column (2) and (3) raises the question of whether the regressions have enough power to explain the variation in adoption of sunflower and sugarcane. If the effects of the network variables are not significantly different from zero simply due to large standard errors caused by smaller sample sizes then the placebo test has no bite. Running the regressions without fixed effects can serve a double purpose: First, it allows me to keep the full sample in the regression to obtain smaller standard errors. Second, it allows me to calculate marginal effects and thereby assess the magnitude of the network effects found on the adoption of placebo crops. However, exploiting variation across subvillages may bias the network estimates if bananas and placebo crops require the same growing conditions which are correlated within subvillages, and if farmers generally discuss farming issues with their neighbors. Table A.4 in Appendix shows the logit estimates and the marginal effects for the network variables when subvillage fixed effects are excluded. In this specification, the adoption of sunflower is significantly correlated with the RIPAT network and the network size, however the marginal effects are miniscule compared to those presented in Table 2. I conclude that the large network effects found cannot be explained by a correlation in entrepreneurship within the networks of farmers who adopt banana cultivation.

Entrepreneurship may not be the only unobserved factor which is correlated within networks. It seems plausible that farmers who prefer a certain crop are more prone to discuss farming issues with each other. Preferences for banana cultivation are most likely captured by the indicator for traditional banana cultivation, but what about other crops? If improved banana cultivation dominates the cultivation of beans, say, then the adoption behavior among bean growers could be correlated, and this would cause a spurious correlation between adoption of banana cultivation and adoption in the network. I can investigate this for the farmers who adopt in the second half of 2009 or later as I have data on the crops grown in 2009. I include indicator variables for the four most popular crops (improved maize, traditional maize, beans and vegetables) to see if the cultivation of any of these crops in 2009 is driving the subsequent adoption and hence the network effects found. The estimates are shown in Table A.5 in Appendix. The effects of the RIPAT and non-RIPAT banana network on adoption are unaffected when I control for the cultivation of improved and traditional maize, beans and vegetables, respectively.[39] When subvillage fixed effects are included, the coefficients to these four cultivation indicators are jointly insignificant.

This further supports the existence of network effects on adoption behavior since it does not provide support for a potential correlation in unobserved farmer characteristics within networks as the main driver of the network effects found.

[39] Column (1) shows that the network effects for this subsample are slightly smaller compared to the full sample estimates from Table 2, before controlling for the cultivation of the other crops.

6.3 Self-selection into RIPAT

The self-selection into RIPAT creates an additional concern. A farmer who knew many farmers who signed up for RIPAT could have chosen not to sign up simply because she knew that she would learn about the new technologies anyway. Since participation in RIPAT required weekly participation in meetings and joint cultivation of the demonstration plot and hence, many work hours, it is a reasonable concern that some farmers who were initially interested in improved banana cultivation could have chosen not to sign up for RIPAT, because several of their network members had done so. This corresponds to the idea of strategic delay derived from the target input model where a farmer would choose to postpone adoption if she knows sufficiently many adopters allowing her to learn from their experimentation without incurring the cost of experimenting herself (Bandiera & Rasul, 2006). Similarly, a farmer could avoid the opportunity cost of labor related to RIPAT participation if one or more network members had chosen to participate from whom she could get improved banana seedlings and instructions.

If I assume that these farmers would adopt banana cultivation relatively early since they were interested in banana cultivation already at the start of the project, I can split the sample of adopters into early and late adopters to see if there are differential effects. If the network effects only persist among early adopters, they may simply be generated by self-selection mechanisms into RIPAT.

I split the sample of adopting farmers on early and late adopters, where 'early adopters' planted their first improved bananas in 2006 or 2007. Within these two years RIPAT farmers would have time to plant bananas on their own farm and the plants would grow sufficiently to produce seedlings that can be shared in the network. I assume that the type of farmers who self-select out of RIPAT because they have network members in a RIPAT group would adopt as soon as possible and hence they would fall in the category of early adopters. If they do believe initially that banana cultivation is more profitable than other crops they grow, the optimal strategy would be to adopt as soon as possible.

Table 8 show the logit estimates with and without farmer and household characteristics and subvillage fixed effects for the two subsamples. There is indeed a larger parameter estimate for the RIPAT network for early adopters but the parameter estimate for the late adopters is very close to the full sample estimate and still highly significant. So even though some of the network effects found in this paper may be explained by self-selection out of RIPAT groups, it appears that social learning also takes place through the perception of the relative profitability of banana cultivation.

7 Conclusion

This paper studies how networks can relax constraints to adoption of a new agricultural technology. The existing literature on networks and adoption of technologies focus on the provision of information through networks (Conley & Udry, 2010; Bandiera & Rasul, 2006; Munshi, 2004; Foster & Rosenzweig, 1995). I contribute to this literature by showing that networks can affect the adoption of a new crop not only through information provision, but also by providing necessary inputs for adoption.

I set up a theoretical model to illustrate how a farmer's network can impact the adoption decision through the two different channels. The model has equivalent implications for both information and input provision through networks which calls for caution when interpreting empirical estimates of network effects.

Empirically, I study the adoption of improved banana cultivation which has been introduced by a project called RIPAT in Tanzania. I study how the adoption among non-RIPAT farmers is affected by their self-reported links to RIPAT farmers and find that knowing at least one banana growing RIPAT farmer increases the propensity to adopt by 37 percentage points. I carefully investigate whether I can consider this estimate to be a causal network effect and I find no evidence for contextual or correlated effects confounding the network estimate. The estimated network effect is most likely a compound effect of information and input provision through the network. Though I cannot fully determine their separate channels, I document that the input provision channel is playing an important role. A solidarity chain principle imbedded in RIPAT obliges participants to pass on thrice as many improved banana seedlings as they have received, and I find that 65 percent of the adopting farmers who knows a RIPAT farmer have received their first improved banana seedlings from him or her.

I furthermore add to the literature on networks and adoption by extending the typical measure of the egocentric network to also include network members who are not growing the crop of interest. The estimates show that they have a negative impact on the adoption decision. The model provides a theoretical explanation for this finding: Network members growing other crops provide information or inputs that makes other crops more attractive, reducing the relative profitability of the crop of interest. For a given amount of land, the farmer is then less likely to adopt the crop of interest. When the total network size is not controlled for in an adoption regression, the effect of the network members who grow the crop of interest may be confounded.

Diffusion of knowledge in networks has received a lot of attention in the literature and recent work by Banerjee et al. (2014) explores the policy question of whom to target to increase the diffusion of knowledge. This is relevant for societies where information flows are hampered by limited access to information technologies.

However, lack of access to information is not the only barrier to adoption of agricultural technologies. In societies with poor infrastructure input markets suffer from high transportation costs which can be a barrier to adoption even if the gross return is high (Suri, 2011). Hence, diffusion of agricultural inputs in networks is a highly relevant topic to study in that context. To my knowledge, Emerick (2013) provides the only empirical study of network trading of agricultural inputs. He finds that input provision through networks is inefficient compared to door-to-door sales of inputs as door-to-door sales lead to a larger degree of adoption in the context of a new rice variety in India. However, the low road density in Sub-Saharan Africa may impede such a market based input distribution due to high transportation costs. The first best solution may be to increase the road density, but according to Spencer (1996) the corresponding costs are so high that it is a more viable strategy to develop agricultural technologies that rely on local input provision.[40]

This is one of the motivations for the solidarity chain principle implemented in RIPAT: When farmers are obliged to pass on improved banana seedlings it creates a local supply of inputs that are necessary for adoption of the new technology. This study documents that the project has successfully fostered diffusion of improved banana cultivation, and that input provision through networks has played an important role. However, the design of the project does not allow me to assess the contribution of the solidarity chain principle to the diffusion of technology. How to best design agricultural projects to foster diffusion of new technologies remains an interesting topic for future research. Not only diffusion of knowledge, but also access to inputs must be addressed, and their separate contributions to the diffusion of new agricultural technologies should be assessed. More research is needed in order to understand when and how network based approaches to the distribution of new inputs can improve on the existing distribution systems.

8 Bibliography

Bandiera, O. & Rasul, I. (2006). Social networks and technology adoption in northern mozambique. *The Economic Journal*, 116, 869–902.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2014). Gossip: Identifying central individuals in a social network. NBER Working Paper No. 20422, National Bureau of Economics Research.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797–817.

Bardhan, P. & Udry, C. (1999). *Development Microeconomics*. Oxford University Press.

Besley, T. & Case, A. (1993). Modeling technology adoption in developing countries. *The American Economic Review*, 83(2), pp. 396–402.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1), 41 – 55.

Brock, W. A. & Durlauf, S. N. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics*, 140(1), 52 – 75.

[40]The efficiency of road construction may have improved since the 1990s but poor infrastructure remains an important challenge in Sub-Saharan Africa (Carruthers et al., 2009).

Cai, J., De Janvry, A., & Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2), 81–108.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414–427.

Carruthers, R., Krishnamani, R. R., & Murray, S. (2009). *Improving Connectivity: Investing in Transport Infrastructure in Sub-Saharan Africa*. Africa infrastructure country diagnostic, World Bank.

Carter, M. R., Laajaj, R., & Yang, D. (2014). Subsidies and the persistence of technology adoption: Field experimental evidence from mozambique. NBER Working Paper No. 20465, National Bureau of Economic Research.

Chamberlain, G. (1982). *Analysis of Covariance With Qualitative Data*. Working Paper 325, National Bureau of Economic Research.

Christiaensen, L., Demery, L., & Kuhl, J. (2011). The (evolving) role of agriculture in poverty reduction–an empirical perspective. *Journal of Development Economics*, 96(2), 239 – 254.

Conley, T. G. & Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review*, 100(1), 35–69.

Davis, K., Nkonya, E., Kato, E., Mekonnen, D., Odendo, M., Miiro, R., & Nkuba, J. (2012). Impact of farmer field schools on agricultural productivity and poverty in east africa. *World Development*, 40(2), 402–413.

de Janvry, A. & Sadoulet, E. (2010). Agricultural growth and poverty reduction: Additional evidence. *The World Bank Research Observer*, 25(1), 1–20.

Diao, X., Hazell, P., & Thurlow, J. (2010). The role of agriculture in african development. *World Development*, 38(10), 1375 – 1383.

Emerick, K. (2013). *The efficiency of trading in social networks: Experimental measures from India*. Job market paper, Department of Agricultural and Resource Economics, University of California Berkeley.

Evenson, R. E. & Westphal, L. E. (1995). Chapter 37 technological change and technology strategy. In J. Behrman & T. Srinivasan (Eds.), *Handbook of Development Economics*, volume 3, Part A (pp. 2209 – 2299). Elsevier.

Feder, G., Just, R. E., & Zilberman, D. (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change*, 33(2), 255–298.

Foster, A. D. & 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.

Foster, A. D. & Rosenzweig, M. R. (2010). Microeconomics of technology adoption. *Annual Review of Economics*, 2(1), 395–424.

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). Agricultural productivity differences across countries. *The American Economic Review*, 104(5), 165–170.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, 25(4), 501–522.

Kremer, M. & Miguel, E. (2007). The illusion of sustainability. *The Quarterly Journal of Economics*, 122(3), 1007–1065.

Krishnan, P. & Patnam, M. (2014). Neighbors and extension agents in ethiopia: Who matters more for technology adoption? *American Journal of Agricultural Economics*, 96(1), 308–327.

Larsen, A. (2012). Adoption of banana cultivation and information networks: An empirical study of northern rural Tanzania. Master's thesis, Department of Economics, University of Copenhagen.

Larsen, A. F. & Lilleør, H. B. (2014). Beyond the field: The impact of farmer field schools on food security and poverty alleviation. *World Development*, 64, 843–859.

Ligon, E. A. & Sadoulet, E. (2011). *Estimating the effects of aggregate agricultural growth on the distribution of expenditures.* CUDARE Working Paper 1115, Department of Agricultural & Resource Economics, UC Berkeley.

Lybbert, T. J. & Sumner, D. A. (2012). Agricultural technologies for climate change in developing countries: Policy options for innovation and technology diffusion. *Food Policy*, 37(1), 114 – 123.

Maertens, A. & Barrett, C. B. (2013). Measuring social networks' effects on agricultural technology adoption. *American Journal of Agricultural Economics*, 95(2), 353–359.

Magnan, N., Spielman, D. J., Gulati, K., & Lybbert, T. J. (2015). *Information Networks among Women and Men and the Demand for an Agricultural Technology in India.* IFPRI discussion paper 01411, International Food Policy Research Institute (IFPRI).

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3), 531–542.

Manski, C. F. (2000). Economic analysis of social interactions. *The Journal of Economics Perspectives*, 14(3), 115–136.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of Econometrics* (pp. 105–142). Academic Press.

Mogensen, H. O. & Pedersen, E. K. (2013). Household dynamics and gender politics: Female farmers in ripat 1. In H. B. Lilleør & U. Lund-Sørensen (Eds.), *Farmers' Choice: Evaluating an approach to agricultural technology adoption*. Practical Action Publishing.

Moser, C. M. & Barrett, C. B. (2006). The complex dynamics of smallholder technology adoption: the case of sri in madagascar. *Agricultural Economics*, 35(3), 373–388.

Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the indian green revolution. *Journal of Development Economics*, 73(1), 185 – 213.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1/2), 122–136.

Prentice, R. L. & Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika*, 66(3), 403–411.

Saha, A., Love, H. A., & Schwart, R. (1994). Adoption of emerging technologies under output uncertainty. *American Journal of Agricultural Economics*, 76(4), 836–846.

Sargent, T. J. (1979). *Macroeconomic Theory*. Academic Press New York.

Schreiner, M. (2011). *A Simple Poverty Scorecard for Tanzania*. Technical report, Microfinance Risk Management. Available online at http://progressoutofpoverty.org, accessed 6 May, 2015.

Shiferaw, B. A., Kebede, T. A., & You, L. (2008). Technology adoption under seed access constraints and the economic impacts of improved pigeonpea varieties in tanzania. *Agricultural Economics*, 39(3), 309–323.

Spencer, D. S. (1996). Infrastructure and technology constraints to agricultural development in the humid and subhumid tropics of africa. *African Development Review*, 8(2), 68–93.

Sunding, D. & Zilberman, D. (2001). Chapter 4 the agricultural innovation process: Research and technology adoption in a changing agricultural sector. In B. L. Gardner & G. C. Rausser (Eds.), *Agricultural Production*, volume 1, Part A of *Handbook of Agricultural Economics* (pp. 207 – 261). Elsevier.

Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, 79(1), 159–209.

Van den Broeck, K. & Dercon, S. (2011). Information flows and social externalities in a tanzanian banana growing village. *The Journal of Development Studies*, 47(2), 231–252.

Waddington, H., White, H., & Anderson, J. (2014). Farmer field schools: From agricultural extension to adult education. *Systematic Review Summary*, 1.

Figures

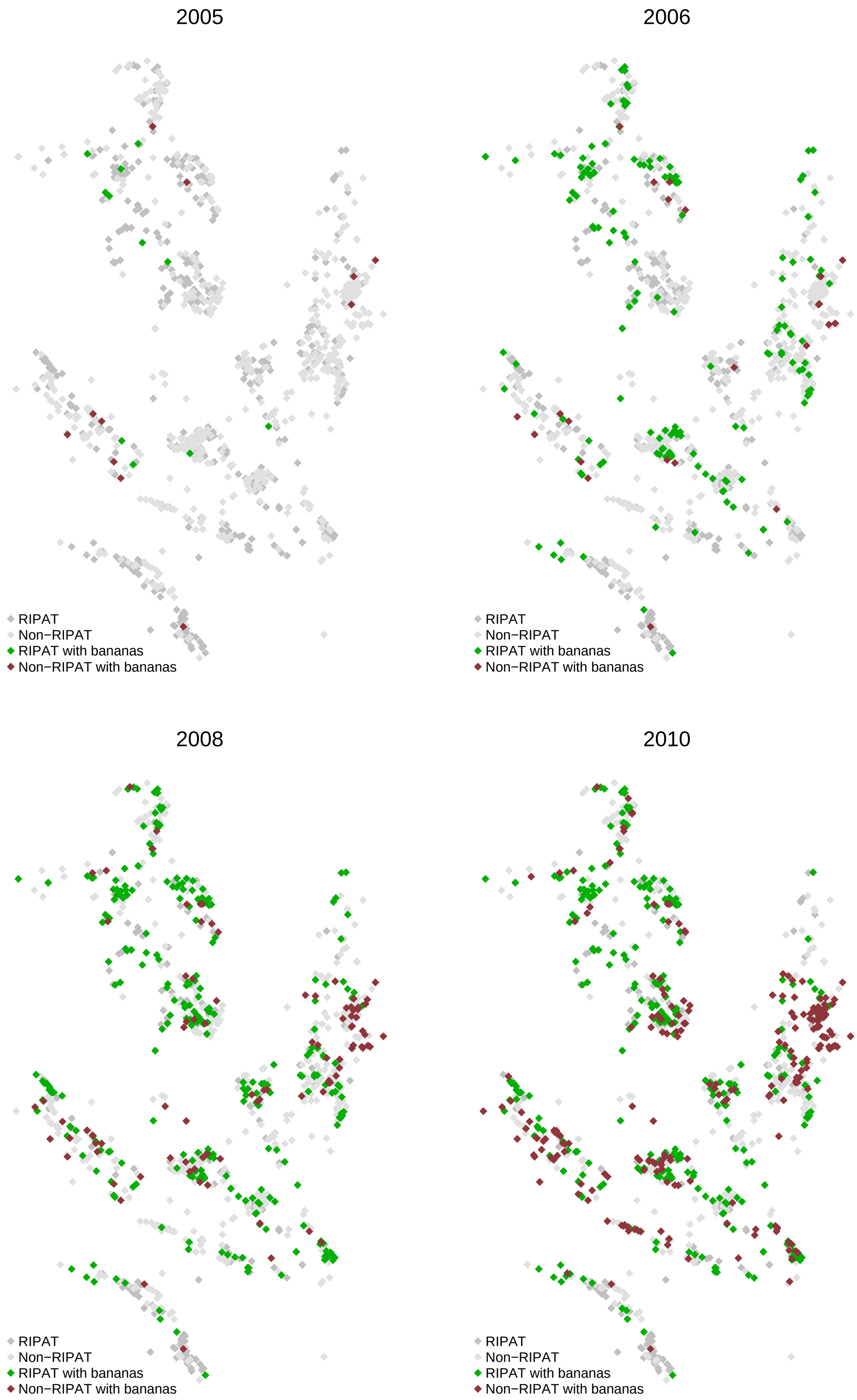
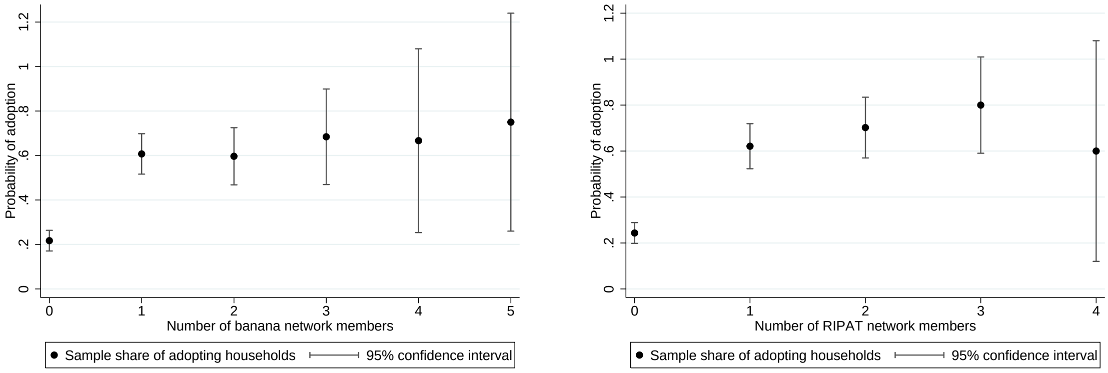


Figure 1: Diffusion of improved banana cultivation



(a) Banana network

(b) RIPAT network

Figure 2: Network measures and sample share of adopting households

Tables

Table 1: Summary statistics

	All		Adopting		Non-adopting		P-value
Grew improved bananas in 2010	0.38	(0.49)	1.00		0.00		
Network variables							
RIPAT banana growers in network	0.52	(0.91)	0.95	(1.10)	0.26	(0.64)	0.000
Non-RIPAT banana growers in network	0.32	(1.87)	0.52	(2.73)	0.20	(1.04)	0.056
Network size	2.83	(4.12)	2.95	(3.79)	2.76	(4.32)	0.613
Farmer characteristics							
Farmer is female	0.22	(0.41)	0.23	(0.42)	0.22	(0.41)	0.736
Age of farmer	44.77	(15.54)	44.06	(13.31)	45.20	(16.76)	0.420
Farmer is Catholic	0.08	(0.28)	0.11	(0.32)	0.07	(0.25)	0.062
Farmer is Muslim	0.05	(0.22)	0.03	(0.16)	0.07	(0.25)	0.033
Farmer has other religion	0.25	(0.43)	0.25	(0.44)	0.25	(0.43)	0.859
Farmer can read	0.20	(0.40)	0.18	(0.38)	0.22	(0.41)	0.287
Household characteristics							
Highest education, less than primary	0.07	(0.25)	0.05	(0.22)	0.08	(0.27)	0.238
Highest education, more than primary	0.35	(0.48)	0.44	(0.50)	0.30	(0.46)	0.002
Household labor	2.54	(1.51)	2.84	(1.63)	2.35	(1.40)	0.000
Household head is widow(er)	0.09	(0.29)	0.06	(0.23)	0.11	(0.31)	0.040
Wealth (poverty score)	44.30	(15.14)	46.32	(14.77)	43.07	(15.26)	0.019
Acres of land 2006	4.21	(6.04)	4.12	(4.36)	4.27	(6.87)	0.780
Distance to nearest road, km	1.45	(1.22)	1.49	(1.26)	1.42	(1.20)	0.526
Participate in other project	0.24	(0.42)	0.28	(0.45)	0.21	(0.41)	0.068
Number of crops grown, 2010	3.96	(1.93)	4.44	(1.95)	3.67	(1.86)	0.000
HH grows/has grown traditional bananas	0.36	(0.48)	0.47	(0.50)	0.29	(0.45)	0.000
No. banana growers within radius of 0.5km	10.83	(8.37)	13.62	(8.61)	9.12	(7.76)	0.000
Observations	509		193		316		

Notes: Sample means and standard deviations in parantheses for all farmers in the sample and for adopting and non-adopting farmers, respectively. The last column presents p-values from double-sided t-tests of equal means for adopting and non-adopting farmers.

Table 2: Adoption of improved banana cultivation

	Logit estimates				Marg.eff.
	(1)	(2)	(3)	(4)	(5)
Network variables					
RIPAT banana growers in network	1.126***	1.105***	0.999***	0.989***	0.241***
	(0.14)	(0.14)	(0.17)	(0.17)	
Non-RIPAT banana growers in network	0.202**	0.210**	0.201**	0.203**	0.049**
	(0.10)	(0.10)	(0.10)	(0.08)	
Network size	-0.124*	-0.127	-0.163**	-0.207***	-0.040**
	(0.08)	(0.08)	(0.08)	(0.06)	
Farmer characteristics					
Farmer is female		0.399	0.592**	0.581*	0.141**
		(0.29)	(0.27)	(0.35)	
Age of farmer		0.058**	0.053	0.043	0.007
		(0.03)	(0.04)	(0.05)	
Age of farmer, sq./100		-0.062**	-0.064*	-0.053	
		(0.03)	(0.04)	(0.05)	
Farmer is Catholic		0.835**	0.817**	0.880*	0.185**
		(0.41)	(0.39)	(0.45)	
Farmer is Muslim		-0.788	-0.211	-0.060	-0.052
		(0.50)	(0.53)	(0.61)	
Farmer has other religion		0.006	-0.055	-0.044	-0.014
		(0.22)	(0.28)	(0.31)	
Farmer can read		-0.092	0.147	0.309	0.036
		(0.29)	(0.33)	(0.39)	
Household characteristics					
Highest education, less than primary			0.218	-0.109	0.053
			(0.35)	(0.66)	
Highest education, more than primary			0.382	0.411	0.093
			(0.26)	(0.29)	
Household labor			0.132	0.188*	0.032
			(0.08)	(0.10)	
Household head is widow(er)			-0.739	-0.585	-0.183
			(0.49)	(0.51)	
Wealth (poverty score)			0.011	0.011	0.002
			(0.01)	(0.01)	
Log of acres, 2006			-0.119	-0.127	-0.017
			(0.19)	(0.17)	
Log distance to road			0.016	-0.103	0.002
			(0.11)	(0.14)	
Participate in NGO project			-0.057	0.113	-0.014
			(0.24)	(0.29)	
Number of crops grown, 2010			0.176**	0.204***	0.044**
			(0.07)	(0.07)	
HH grows/has grown traditional bananas			0.689**	0.625**	0.165**
			(0.28)	(0.26)	
No. banana growers within radius of 0.5km			0.055***	0.033*	0.008***
			(0.02)	(0.02)	
Subvillage fixed effects	No	No	No	Yes	No
Observations	509	509	509	503	509

Notes: The dependent variable is an indicator equal to one if the farmer grows improved bananas in 2010. Logit coefficient estimates are presented in column (1)-(4), constant not reported. Standard errors in parentheses are clustered at the subvillage level. Marginal effects presented in column (5) are calculated as described in the text based on estimates from column (3). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3: Extensive and intensive margin network

	Logit estimates				Marg.eff.
	(1)	(2)	(3)	(4)	(5)
1+ RIPAT banana grower in network	1.789***	1.760***	1.593***	1.418***	0.365***
	(0.31)	(0.31)	(0.37)	(0.30)	
2+ RIPAT banana growers in network	0.494	0.488	0.664	0.834*	0.096
	(0.43)	(0.43)	(0.46)	(0.47)	
3+ RIPAT banana growers in network	0.279	0.284	-0.260	-0.320	-0.033
	(0.48)	(0.49)	(0.52)	(0.74)	
1+ non-RIPAT banana grower in network	1.073***	1.104**	1.216***	0.667*	0.266***
	(0.41)	(0.44)	(0.42)	(0.40)	
2+ non-RIPAT banana growers in network	-0.475	-0.489	-0.908	-0.583	-0.190
	(0.67)	(0.69)	(0.83)	(0.89)	
3+ non-RIPAT banana growers in network	0.911	1.012	1.242	1.186	0.241
	(1.13)	(1.09)	(1.30)	(1.24)	
Network size	-0.089*	-0.090*	-0.124**	-0.137***	-0.030**
	(0.05)	(0.05)	(0.06)	(0.05)	
Farmer characteristics	No	Yes	Yes	Yes	Yes
Household characteristics	No	No	Yes	Yes	Yes
Subvillage fixed effects	No	No	No	Yes	No
Observations	509	509	509	503	509

Notes: The dependent variable is an indicator equal to one if the farmer grows improved bananas in 2010. Logit coefficient estimates are presented in column (1)-(4), standard errors in parentheses are clustered at the subvillage level. Column (5) presents marginal effects calculated as described in the text. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 4: Where did the adopting farmers get the first banana seedlings?

	Received seedling from					
Sample	RIPAT in network	RIPAT not in network	Non-RIPAT	Other	Total	Obs.
All adopting farmers	36.9%	29.4%	25.3%	8.4%	100%	193
Know 1+ RIPAT banana grower	65.4%	14.0%	15.9%	4.7%	100%	107
Know 1+ non-RIPAT banana grower	25.7%	25.7%	37.1%	11.5%	100%	35
Know both	52.9%	11.8%	29.4%	5.9%	100%	17

Notes: The numbers in the first five columns are percentages while the last column gives the number of observations for the given row. Percentages in each row sum to 100. The category "Other" covers the implementing organization, a RIPAT group, or another NGO. The last row includes adopting farmers who discuss farming issues with at least one RIPAT banana grower *and* at least one non-RIPAT banana grower.

Table 5: Reasons for not adopting improved bananas

Why have you not planted any improved bananas?	mean	std.err.
I do not know the production techniques	0.133	(0.02)
My soil is inadequate for banana cultivation	0.041	(0.01)
I have never grown bananas	0.044	(0.01)
The work is too hard/tough	0.101	(0.02)
I don't have enough land	0.174	(0.02)
I think it will not be remunerative	0.006	(0.00)
I prefer growing traditional bananas	0.032	(0.01)
Weather is not suitable	0.057	(0.01)
Water shortage	0.392	(0.03)
Never got seedlings	0.130	(0.02)
Other	0.006	(0.00)
Observations	316	

Notes: Multiple answers were possible so the shares do not sum to one.

Table 6: Adoption of improved banana cultivation, split on characteristics of RIPAT network

	Wealth (1)	Land (2)	Education (3)	Gender (4)	Age (5)
Poor RIPAT banana growers	1.144***				
	(0.25)				
Rich RIPAT banana growers	0.842***				
	(0.23)				
RIPAT banana growers w. small farm		1.255***			
		(0.24)			
RIPAT banana growers w. large farm		0.719***			
		(0.24)			
RIPAT banana growers with low edu.			0.474		
			(0.36)		
RIPAT banana growers with medium edu.			1.184***		
			(0.22)		
RIPAT banana growers with high edu.			0.926		
			(0.75)		
Male RIPAT banana growers				1.066***	
				(0.20)	
Female RIPAT banana growers				0.764**	
				(0.31)	
Young RIPAT banana growers					1.000***
					(0.26)
Old RIPAT banana growers					1.009***
					(0.26)
Non-RIPAT banana growers in network	0.205**	0.215**	0.208**	0.206**	0.214**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Network size	-0.204***	-0.195***	-0.210***	-0.198***	-0.204***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Observations	503	503	503	503	503
χ^2 test of equality[a]	0.84	2.56	3.00	0.71	0.00
P-value	0.360	0.109	0.223	0.398	0.981

Notes: The dependent variable is an indicator equal to one if the farmer grows improved bananas in 2010. Conditional logit coefficient estimates controlling for farmer and household characteristics and accounting for subvillage fixed effects. Standard errors in parentheses are clustered at the subvillage level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

[a] The following is tested: Column (1): Poor = rich (df = 1), Column (2): small = large (df = 1), Column (3): Low edu. = medium edu. = high edu. (df = 2), Column (4): Male = female (df = 1), Column (5): Young = old (df = 1).

Table 7: Placebo study results, adoption of vegetables, sunflowers and sugarcanes

	(1) Vegetables	(2) Sunflower	(3) Sugarcane
RIPAT banana growers in network	0.261	0.222	-0.030
	(0.26)	(0.20)	(0.25)
Non-RIPAT banana growers in network	0.139	-0.003	-0.064
	(0.19)	(0.15)	(0.19)
Network size	0.047	0.024	-0.031
	(0.03)	(0.08)	(0.10)
RIPAT growing vegetables	-0.169		
	(0.34)		
RIPAT growing sunflowers		0.295	
		(0.58)	
RIPAT growing sugarcanes			0.763
			(0.97)
Observations	487	337	305
Mean of dependent variable	0.489	0.134	0.081
Std.dev. of dependent variable	(0.500)	(0.341)	(0.272)

Notes: The dependent variables are a indicators equal to one if the farmer grows vegetables (column 1), sunflowers (column 2) and sugarcanes (column 3) in 2010. Conditional logit coefficient estimates accounting for subvillage fixed effects. Standard errors in parentheses are clustered at the subvillage level. Farmer and household characteristics are included in all specifications, but in column (2) religion dummies are excluded as Catholic dummy predicts non-adoption perfectly leading to drop of 43 observations. Results are robust to inclusion of religion dummies. The number of crops grown in 2010 is subtracted traditional and improved bananas and the placebo crop. Due to lack of variation in adoption within some subvillages, the number of observations is lower than 509.

Table 8: Adoption of banana cultivation, early and late adopter subsamples

	Early adopters			Late adopters		
	(1)	(2)	(3)	(4)	(5)	(6)
RIPAT network	1.497***	1.393***	1.394***	1.049***	0.964***	0.903***
	(0.24)	(0.22)	(0.32)	(0.13)	(0.17)	(0.18)
Non-RIPAT network	0.386**	0.401**	0.318	0.195*	0.186*	0.191**
	(0.19)	(0.18)	(0.30)	(0.10)	(0.10)	(0.09)
Network size	-0.399**	-0.504***	-0.536***	-0.108	-0.136*	-0.180***
	(0.16)	(0.16)	(0.17)	(0.07)	(0.07)	(0.06)
Farmer characteristics	No	Yes	Yes	No	Yes	Yes
Household characteristics	No	Yes	Yes	No	Yes	Yes
Subvillage fixed effects	No	No	Yes	No	No	Yes
Observations	358	358	256	464	464	458

Notes: The dependent variable is an indicator equal to one if the farmer grows improved bananas in 2010. Logit coefficient estimates, standard errors in parentheses clustered at the subvillage level. Column (1)-(3) is based on data from non-adopting households and households who adopted in 2006 or 2007. Column (4)-(6) is based on data non-adopting households and households who adopted in 2008 or later. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Appendix

A Derivation of mean-variance expected utility

To see how the expected utility $E[U(y)] = E\left[-e^{-\lambda((1-\omega)y_a+\omega y_b)}\right]$ can be rewritten to depend on the expected mean and variance of y_b, I first write the expected utility as

$$E[U(y)] = \frac{-e^{-\lambda(1-\omega)y_a}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} -e^{-\lambda\omega y_b} e^{-\frac{(y_b-\mu)^2}{2\sigma^2}} dy_b \tag{A.1}$$

I rewrite the exponent within the integral into two terms where one does not depend on y_b:

$$\lambda\omega y_b + \frac{(y_b-\mu)^2}{2\sigma^2} = \frac{(y_b-\mu+\omega\lambda\sigma^2)^2}{2\sigma^2} + \lambda\left(\mu\omega - \frac{\omega^2\lambda\sigma^2}{2}\right)$$

Inserting this exponent in equation A.1 and rearranging gives

$$E[U(y)] = \frac{-e^{-\lambda((1-\omega)y_a+\mu\omega-\frac{1}{2}\omega^2\lambda\sigma^2)}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y_b-\mu+\omega\lambda\sigma^2)^2}{2\sigma^2}} dy_b$$

Now, I exploit that for all $\tilde{\mu}$ (including $\tilde{\mu} = \mu - \omega\lambda\sigma^2$)

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y_b-\tilde{\mu})^2}{2\sigma^2}} dy_b = 1$$

because the left hand side is just the total area under the density function when the mean is $\tilde{\mu}$ and the standard deviation is σ^2. This implies that the expression for the expected utility simplifies to

$$E[U(y)] = -e^{-\lambda((1-\omega)y_a+\mu\omega-\frac{1}{2}\omega^2\lambda\sigma^2)}$$

B Choice-based sampling in a logit model

This section shows that the logit model provides consistent estimates of the parameters in the case of choice-based sampling.[41]

Assume that the probability of adoption in the population, $\tilde{P}(a_i = 1)$, is logistically distributed and depends on a range of farmer and household characteristics, Z_i and subvillage fixed effects, α_s:

$$\tilde{P}(a_i = 1|Z_i, \alpha_s) = \Lambda(\theta Z_i + \alpha_s) = \frac{exp(\theta Z_i + \alpha_s)}{1 + exp(\theta Z_i + \alpha_s)} \tag{B.1}$$

For simplicity assume that all the covariates are discrete, such that I can consider probabilities instead of distributions. The result generalizes to continuous covariates.

[41] I would like to thank Professor Bo Honoré, Princeton University, for indispensable help with the following derivation.

The probability of adoption in the sample, $P(a_i = 1)$, conditional on covariates and subvillage fixed effects can be rewritten using Bayes rule:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{P(a_i = 1, Z_i, \alpha_s)}{P(Z_i, \alpha_s)}$$

I now use that the sample of adopting farmers is a random sample such that the probability of the covariates given that the farmer is adopting is the same in the sample and in the population, $P(Z_i, \alpha_s|a_i = 1) = \tilde{P}(Z_i, \alpha_s|a_i = 1)$, and correspondingly for non-adopting farmers. In addition, I use the law of iterated expectations in the denominator:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{\tilde{P}(Z_i, \alpha_s|a_i = 1) \cdot P(a_i = 1)}{\sum \tilde{P}(Z_i, \alpha_s|a_i = 0) \cdot P(a_i = 0) + \sum \tilde{P}(Z_i, \alpha_s|a_i = 1) \cdot P(a_i = 1)} \tag{B.2}$$

Applying Bayes rule and using equation B.1, I can rewrite:

$$\tilde{P}(Z_i, \alpha_s|a_i = 1) = \frac{\tilde{P}(Z_i, \alpha_s, a_i = 1)}{\tilde{P}(a_i = 1)} = \frac{\tilde{P}(Z_i, \alpha_s) \cdot \Lambda(\theta Z_i + \alpha_s)}{\tilde{P}(a_i = 1)} \tag{B.3}$$

Correspondingly,

$$\tilde{P}(Z_i, \alpha_s|a_i = 0) = \frac{\tilde{P}(Z_i, \alpha_s) \cdot (1 - \Lambda(\theta Z_i + \alpha_s))}{\tilde{P}(a_i = 0)} \tag{B.4}$$

I now insert equation B.3 and B.4 in equation B.2:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{\frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \tilde{P}(Z_i, \alpha_s) \cdot \Lambda(\theta Z_i + \alpha_s)}{\frac{P(a_i=0)}{\tilde{P}(a_i=0)} \tilde{P}(Z_i, \alpha_s) \cdot (1 - \Lambda(\theta Z_i + \alpha_s)) + \frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \tilde{P}(Z_i, \alpha_s) \cdot \Lambda(\theta Z_i + \alpha_s)}$$

I divide numerator and denominator with $\tilde{P}(Z_i, \alpha_s)$ and insert the definition of the logistic distribution:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{\frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \frac{exp(\theta Z_i + \alpha_s)}{1 + exp(\theta Z_i + \alpha_s)}}{\frac{P(a_i=0)}{\tilde{P}(a_i=0)} \cdot \frac{1}{1 + exp(\theta Z_i + \alpha_s)} + \frac{P(a_i=1)}{\tilde{P}(a_i=1)} \cdot \frac{exp(\theta Z_i + \alpha_s)}{1 + exp(\theta Z_i + \alpha_s)}}$$

Finally, I divide both numerator and denominator with the first term of the denominator and rearrange:

$$P(a_i = 1|Z_i, \alpha_s) = \frac{exp(\theta Z_i + \alpha_s + ln(c))}{1 + exp(\theta Z_i + \alpha_s + ln(c))}, \qquad c \equiv \frac{P(a_i=1)/\tilde{P}(a_i=1)}{P(a_i=0)/\tilde{P}(a_i=0)} \tag{B.5}$$

Comparing the probability of adoption in the sample (equation B.5) with the probability of adoption in the population (equation B.1), it is evident that the choice-based sampling only affects the estimation of the subvillage fixed effects (or the constant in the case of no fixed effects) and hence, the estimated parameters of the covariates (θ) are unaffected by the sampling method.

C Correction of the constant term for calculation of marginal effects

In order to calculate marginal effects I need to correct the constant term with the factor $ln(c)$.

$$
\begin{aligned}
ln(c) &= ln\left(\frac{P(a_i=1)/\tilde{P}(a_i=1)}{P(a_i=0)/\tilde{P}(a_i=0)}\right) \\
&= ln\left(P(a_i=1)\right) + ln\left(\tilde{P}(a_i=0)\right) - ln\left(\tilde{P}(a_i=1)\right) - ln\left(P(a_i=0)\right)
\end{aligned}
$$

The population and sample probability differ due to non-random sampling. I will simply estimate the sample probabilities by sample proportions:

- $P(a_i = 1)$ is equal to the share of adopting farmers in the final sample of nonRIPAT farmers
- $P(a_i = 0)$ is equal to the share of non-adopting farmers in the final sample of nonRIPAT farmers

The population probabilities are more complicated to calculate due to a complex sampling design. It is briefly described in the following, for more details see Appendix A in Larsen (2012).

We drew a random sample of size R in each village among which the village leader identified the adopting households. Let L_i be the indicator for being identified as adopting household by the village leader. This is however not a perfect measure of adoption, $P(L_i = 1) \neq P(a_i = 1)$. Within the random sample we only interviewed a stratified subsample, which was stratified to achieve the same share of adopting and non-adopting farmers based on village leader identification. The population probability of being an adopting household can be calculated as:

$$\tilde{P}\left(a_i = 1\right) = P\left(a_i = 1|L_i = 1\right) \cdot P\left(L_i = 1\right) + P\left(a_i = 1|L_i = 0\right) \cdot P\left(L_i = 0\right) \tag{C.1}$$

All elements can be estimated by sample shares:

- $\hat{P}\left(a_i = 1|L_i = 1\right)$: the share of adopting households in the final random sample among those identified as adopting by the village leader
- $\hat{P}\left(a_i = 1|L_i = 0\right)$: the share of adopting households in the final random sample among those identified as non-adopting by the village leader
- $\hat{P}\left(L_i = 1\right)$: the share of adopting households as identified by the village leader in the random sample
- $\hat{P}\left(L_i = 0\right)$: the share of non-adopting households as identified by the village leader in the random sample

These sample shares are then inserted in equation C.1 to calculate the population probability of being an adopting household. The probability of being a non-adopting household can be calculated in a corresponding way.

D Appendix tables

Table A.1: OLS results with wild bootstrap-t p-values in square brackets

	(1)	(2)	(3)
PANEL A: Linear network variables			
RIPAT banana growers in network	0.192	0.192	0.149
	(0.030)	(0.031)	(0.028)
	[0.000]	[0.000]	[0.000]
Non-RIPAT banana growers in network	0.041	0.042	0.026
	(0.015)	(0.015)	(0.015)
	[0.038]	[0.030]	[0.302]
Network size	-0.024	-0.024	-0.023
	(0.008)	(0.008)	(0.007)
	[0.002]	[0.002]	[0.002]
PANEL B: Extensive and intensive margin			
1+ RIPAT banana grower in network	0.255	0.253	0.195
	(0.077)	(0.078)	(0.079)
	[0.008]	[0.010]	[0.044]
2+ RIPAT banana growers in network	0.202	0.205	0.191
	(0.137)	(0.142)	(0.125)
3+ RIPAT banana growers in network	-0.034	-0.030	-0.103
	(0.142)	(0.144)	(0.146)
1+ non-RIPAT banana grower network	0.153	0.158	0.137
	(0.096)	(0.096)	(0.086)
	[0.284]	[0.264]	[0.346]
2+ non-RIPAT banana growers network	0.073	0.069	-0.052
	(0.127)	(0.131)	(0.119)
3+ non-RIPAT banana growers network	-0.035	-0.028	0.022
	(0.147)	(0.149)	(0.147)
Network size	-0.022	-0.022	-0.023
	(0.008)	(0.008)	(0.007)
	[0.002]	[0.002]	[0.002]
Farmer characteristics	No	Yes	Yes
Household characteristics	No	No	Yes
Observations	509	509	509

Notes: The dependent variable is an indicator equal to one if the farmer grows improved bananas in 2010. OLS estimates, standard errors in parentheses clustered at the subvillage level, wild bootstrap-t p-values are presented in square brackets calculated as suggested by Cameron et al. (2008). Observations are weighted with inverse sampling probability weights.

Table A.2: Network estimates using random sample only

	Logit estimates				Marg.eff.
	(1)	(2)	(3)	(4)	(5)
Panel A: Main specification					
RIPAT banana growers in network	1.481***	1.546***	1.298***	1.291***	0.279***
	(0.22)	(0.23)	(0.27)	(0.32)	
Non-RIPAT banana growers in network	0.356**	0.347**	0.273*	0.186	0.056*
	(0.15)	(0.14)	(0.16)	(0.22)	
Network size	-0.358***	-0.365***	-0.392***	-0.397***	-0.078***
	(0.14)	(0.13)	(0.14)	(0.11)	
Panel B: Extensive and intensive margin					
1+ RIPAT banana grower in network	2.208***	2.307***	2.018***	1.623***	0.442***
	(0.49)	(0.46)	(0.55)	(0.46)	
2+ RIPAT banana growers in network	0.680	0.763	0.736	1.114	0.154
	(0.72)	(0.80)	(0.76)	(0.80)	
3+ RIPAT banana growers in network	0.250	0.362	-0.755	-0.720	-0.158
	(0.95)	(1.04)	(1.25)	(1.54)	
1+ non-RIPAT banana grower in network	1.155**	1.137**	1.237***	0.638	0.279***
	(0.54)	(0.53)	(0.45)	(0.59)	
2+ non-RIPAT banana growers in network	0.479	0.584	-0.062	0.309	-0.016
	(0.78)	(0.91)	(1.23)	(1.18)	
Network size	-0.371***	-0.378***	-0.389***	-0.400***	-0.075***
	(0.13)	(0.13)	(0.14)	(0.11)	
Farmer characteristics	No	Yes	Yes	Yes	Yes
Household characteristics	No	No	Yes	Yes	Yes
Subvillage fixed effects	No	No	No	Yes	No
Observations	356	356	356	309	356

Notes: The dependent variable is an indicator equal to one if the farmer grows improved bananas in 2010. Only the random sample of farmers is used to produce the estimates. Panel A presents network estimates based on the main specification as presented in table 2. Panel B presents the extensive and intensive margin estimates as presented in table 3, however as only 4 farmers know 3 or more non-RIPAT banana growers in the random sample, this variable is excluded. Logit coefficient estimates are presented in column (1)-(4), standard errors in parentheses are clustered at subvillage level. Marginal effects presented in column (5) are calculated as described in the text based on estimates from column (3). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.3: Having a banana plantation with ten or more plants

	Logit estimates				Marg.eff.
	(1)	(2)	(3)	(4)	(5)
RIPAT banana growers in network	1.093***	1.061***	0.985***	0.881***	0.235***
	(0.15)	(0.16)	(0.19)	(0.17)	
Non-RIPAT banana growers in network	0.272**	0.280**	0.310***	0.308***	0.076***
	(0.11)	(0.11)	(0.10)	(0.10)	
Network size	-0.163	-0.176*	-0.206**	-0.255***	-0.051**
	(0.11)	(0.10)	(0.09)	(0.07)	
Farmer characteristics	No	Yes	Yes	Yes	Yes
Household characteristics	No	No	Yes	Yes	Yes
Subvillage fixed effects	No	No	No	Yes	No
Observations	509	509	509	496	509

Notes: Dependent variable is an indicator variable equal to one if the farmer has a banana plantation with ten banana plants or more. Logit coefficient estimates in column (1)-(4), standard errors in parentheses, clustered at subvillage level. Marginal effects presented in column (5) are calculated as described in the text based on estimates from column (3). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.4: Placebo results without subvillage fixed effects and with marginal effects

	Vegetables		Sunflower		Sugar canes	
	(1) logit	(2) marg.eff.	(3) logit	(4) marg.eff.	(5) logit	(6) marg.eff.
RIPAT banana growers in network	-0.046	-0.010	0.263**	0.034**	0.115	0.008
	(0.25)		(0.10)		(0.14)	
Non-RIPAT banana growers in network	0.131	0.028	-0.036	-0.005	-0.007	-0.000
	(0.15)		(0.09)		(0.08)	
Network size	0.027	0.006	0.049**	0.006**	-0.060	-0.004
	(0.03)		(0.02)		(0.09)	
RIPAT growing vegetables	0.401	0.083				
	(0.33)					
RIPAT growing sunflowers			0.758**	0.123**		
			(0.38)			
RIPAT growing sugarcanes					1.628***	0.216***
					(0.59)	
Observations	509		509		509	
Mean of dependent variable	0.489		0.134		0.081	
Std.dev. of dependent variable	0.500		0.341		0.272	

Notes: The dependent variables are a indicators equal to one if the farmer grows vegetables (column 1-2), sunflowers (column 3-4) and sugarcanes (column 5-6) in 2010. Conditional logit coefficient estimates are presented in column (1), (3) and (5), standard errors in parentheses are clustered at the subvillage level. The remaining columns present marginal effects calculated at the sample mean. Farmer and household characteristics are included in all specifications, but in column (3)-(4) religion dummies are excluded as Catholic dummy predicts non-adoption perfectly leading to drop of 43 observations. The number of crops grown in 2010 is subtracted traditional and improved bananas and the placebo crop. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.5: Adoption of banana cultivation in second half of 2009 or later, controlling for other crops grown in 2009

	Logit estimates					Marg.eff.
	(1)	(2)	(3)	(4)	(5)	(6)
RIPAT banana growers in network	0.783***	0.863***	0.873***	0.855***	0.807***	0.183***
	(0.23)	(0.20)	(0.19)	(0.22)	(0.23)	
Non-RIPAT banana growers in network	0.123	0.134	0.130	0.118	0.128	0.025
	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)	
Network size	-0.120*	-0.076	-0.074	-0.088	-0.129*	-0.018
	(0.07)	(0.09)	(0.08)	(0.09)	(0.07)	
Grows improved maize in 2009		0.233	0.296	0.125	0.109	0.026
		(0.29)	(0.30)	(0.36)	(0.42)	
Grows traditional maize in 2009		-0.060	-0.094	-0.174	0.104	-0.037
		(0.33)	(0.34)	(0.35)	(0.41)	
Grows beans in 2009		0.853**	0.828**	0.929**	0.768	0.168**
		(0.41)	(0.40)	(0.40)	(0.50)	
Grows vegetables in 2009		0.233	0.283	-0.086	-0.243	-0.018
		(0.28)	(0.27)	(0.35)	(0.40)	
Farmer char.	Yes	No	Yes	Yes	Yes	Yes
Household char.	Yes	No	No	Yes	Yes	Yes
Subvillage fixed effects	Yes	No	No	No	Yes	No
Observations	318	392	392	392	318	392
P-value (testing other crops = 0)		0.045	0.015	0.068	0.587	

Notes: The dependent variable is an indicator equal to one if the farmer grows improved bananas in 2010. Farmers who planted bananas before the second half of 2009 are exluded from the sample. Column (1) provides logit estimates using the same specification as column (4) in Table 2 on this subsample. Column (2)-(5) provides logit estimates while column (6) presents marginal effects based on column (4) estimates. Standard errors in parentheses are clustered at the subvillage level. The bottom row presents p-values from χ^2 tests of joint insignificance of the coefficients to the four other crops (improved and traditional maize, beans and vegetables). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.