Learning from Others in Heterogeneous Environments*

Emilia Tjernström
University of Wisconsin, Madison

3rd April 2017

Abstract

This paper examines the influence of social networks on the diffusion of a new agricultural technology that is sensitive to underlying farm characteristics. We show that soil quality heterogeneity affects farmers' ability to learn from their peer's experimentation with the new technology. Modeling the expected output of heterogeneous plots conditional on inputs as a realization of a Brownian motion process, we show that the variance of a farmer's posterior beliefs about their field's returns to the technology increases in the variance of soil quality in the village. Exploiting data from the randomized introduction of a new technology in rural Kenyan villages, which induces experimental variation in the information available to farmers through their social networks, and data on plot-level soil fertility, we show that soil quality heterogeneity indeed weakens the strength of learning-from-others in this context. The results suggest that policies and programs that attempt to leverage social learning to diffuse technologies need to take into account the complex learning process that farmers face.

1 Introduction

Understanding when and how individuals learn from their social connections is especially important in settings where formal institutions are weak. In the presence of market frictions and incomplete information about the availability and profitability of new technologies, the beliefs and actions of one's peer group are likely to act as substitutes for more formal information channels. In the

*An earlier version of this paper was circulated under the title "Signals, Similarity and Seeds: Social Learning in the Presence of Imperfect Information and Heterogeneity." I thank Ivan Balbuzanov, Thomas Barré, Steve Boucher, Phil Brown, Isabel Call, Michael Carter, Andrew Crane-Droesch, Ghada Elabed, Jon Einar Flåtnes, Jason Kerwin, Travis Lybbert, Laura Schechter, Chris Udry, Wenbo Zou, and seminar participants at Bowdoin College, Colby College, Copenhagen University, Georgetown, Ohio State University, Oregon State University, Paris School of Economics, Université Laval, University of California - Davis, University of California - Berkeley, University of Illinois - Urbana-Champaign, University of Minnesota, and University of Wisconsin - Madison for helpful comments and suggestions. This research, commissioned by Acumen, a non-profit impact investment firm, is made possible in part by the generous support of the American people through the United States Agency for International Development Cooperative Agreement No. AID-OAA-L-12-00001 with the BASIS Feed the Future Innovation Lab. I also thank the Agricultural Technology Adoption Initiative (ATAI) administered by JPAL at MIT and the Bill and Melinda Gates Foundation for funding. The contents are my responsibility and do not necessarily reflect the views of USAID, the US Government or other funders. I also wish to thank our respondents for their enthusiasm, generosity and patience during the course of this research.

context of technology diffusion, the extent to which network-based spillovers on technology adoption enhances social welfare depends partly on the source of the spillovers. A conformity motive may result in widespread adoption of a suboptimal technology, while responsiveness to network-based information about the payoffs to a new technology is more likely to generate an advantageous diffusion process.

We exploit a randomized experiment that generated exogenous variation in the information available to individuals through their social network to study the types of information that matter for learning, and the effect of heterogeneity on the learning process. By explicitly eliciting treated farmers' experience with a new, randomly distributed, technology, we can identify how farmers respond to signals received through their network about the returns to the technology, above and beyond the influence of the number of treated links in their network. We further combine the network information with detailed soil quality information and find that heterogeneity in the returns to adoption, in the form of variation in soil fertility within a village, hampers network-driven learning by making it harder for farmers to learn from each other.

Low agricultural productivity in the developing world is a problem at both micro and macro levels: three-quarters of poor people in developing countries live in rural areas and depend at least in part on agriculture for their livelihoods (World Bank, 2008). Further, studies show that GDP growth originating in agriculture benefits the poor substantially more than growth originating in other sectors (Ligon and Sadoulet, 2008). Although technologies exist that could raise agricultural productivity, thereby lifting poor households out of poverty and growing the economies in which they live, these technologies have not been widely adopted in poor countries, especially those in sub-Saharan Africa. Understanding how farmers in these countries make decisions about which technologies to use is therefore a key issue in closing the yield gap and improving the welfare of agricultural households; this paper studies how social networks influence information diffusion and adoption decisions among farmers in Western Kenya.

The study of social learning has grown in popularity over the past few decades and researchers have employed various careful econometric strategies to circumvent the fact that observing similar behavior among members of a social network does not constitute sufficient evidence of social learning. The primary challenge (detailed by Manski, 1993) is identifying whether members of a social network influence each other or whether they simply behave alike because they are already similar and face similar stochastic shocks (perhaps because of a shared environment or because the network was formed precisely based on the shared characteristics of its members). Until recently, social network studies in economics relied exclusively on observational data and typically used panel data coupled with innovative measures of the information available to individuals, or the predictions of theoretical models, to identify network effects. Some examples with agricultural technologies include Conley and Udry (2010), who exploit the timing of plantings and a measure of new/surprising information about input productivity, Munshi (2004), who uses lagged acreage decisions and yield realizations, and Maertens (2014), who employs an aggregate measure of the information available to farmers on current and historical inputs and outputs of other farmers.

Besley and Case (1994) and Foster and Rosenzweig (1995) take more structural approaches to the estimation of learning, leveraging assumptions on steady-state learning and the level of variability in optimal management of the new technology within villages for identification, respectively.

A growing number of studies solve the identification problem by experimentally influencing the information available through social networks (see, for example, Babcock and Hartman, 2010, Carter et al., 2014a, Cai et al., 2015, Magnan et al., 2013, and Oster and Thornton, 2012). By inducing exogenous variation in respondents' social networks, these papers get around the core identification problem and credibly identify the causal effects of social networks on various measures of technology adoption. In contrast to the more nuanced measures of information employed by the observational literature, however, all the above papers base their main social network analyses on how many members of an individual's network were treated, arguing that this is an appropriate measure (see, for example, Bandiera and Rasul, 2006) since it proxies for the number of different sources of information to which a farmer has access.

While this popular approach is useful and has contributed to establishing the importance of network effects in the literature, the implicit adoption mechanism relates more closely to the social influence models used in sociological studies of innovation than to the social learning models common in economics. Social influence models consider social pressure to be a key driver of adoption; individuals are typically assumed to base adoption decisions on the number (or proportion) of adopters in the population or network.[1] We complement this method with more precise measures of the information available through farmers' social networks, echoing the approach in Conley and Udry (2010). Our information measure differs from theirs in that we explicitly elicit farmers' experiences with the technology instead of relying on observed input use. Specifically, some farmers in our sample (the directly treated) received small packs of a new maize variety and conducted on-farm trials with the seeds. Their fellow villagers (the indirectly treated) only have access to information about the seeds through their social networks. We obtained the directly treated farmers' evaluation of how well the on-farm experiment went and let the signal that a given farmer receives about the new technology be a function of the distribution of these evaluations in her information network.

Observing peer effects may reflect mimicry or social pressure rather than actual learning, but these more precise measures of information enable us to more carefully discern between imitation and real learning. Few experimental network studies are able to identify the underlying drivers of network effects. Cai et al. (2015) try to isolate learning effects by eliciting farmers' understanding about how a new weather insurance technology works, thereby directly testing whether information was transmitted and learned through the network. Oster and Thornton (2012) attempt to tease out whether the impact of peers on menstrual cup adoption arises due to learning by others in their network or due to an impact on their friends' valuation of the technology. We propose an alternative way of distinguishing learning from imitation that involves contrasting individuals' responsiveness

[1] See Young (2009) for a detailed description of three separate models (contagion, social influence, and social learning) and a discussion of how to incorporate heterogeneity into these models in a general way.

to the *number of people* who have experience with the new technology with their response to the *actual information* being transmitted through the network. If individuals respond to the number of people in their social network who adopt a new technology, but not to information about the profitability of this technology, then any observed peer effects are more likely to be a sign of mimicry than of social learning.

To the contrary, while the empirical results from our social-influence regressions suggest that networks transmit information and affect respondents' willingness to pay for the hybrids, the indirectly treated farmers respond strongly to the signals available in the network, above and beyond the impact of the number of treated links in their network.[2] In particular, we find that the information signals impact the indirectly treated farmers' familiarity with the new technology, their willingness to pay for the seeds, and – once we control for the impact of heterogeneity – their probability of adopting the new technology. For treated farmers, social networks have only moderate impact on individuals' familiarity with the technology, but they affect farmers' willingness to pay for the new seeds as well as their probability of adoption.

Our second main result takes advantage of detailed measures of the soil quality on treated farmers' plots. We find that the observed social network effects are weaker in villages in which soil quality is more varied, which illustrates how heterogeneity in returns can handicap network effects. These results constitute, to our knowledge, the first empirical evidence on the link between heterogeneity in individual characteristics and information diffusion within the context of a single agricultural technology. Munshi (2004) compares farmers from rice- and wheat-growing regions of India during the Green Revolution to show that social learning is weaker in heterogeneous populations when the new technology is more sensitive to the characteristics in which there is unobserved heterogeneity. However, since the variation in heterogeneity in his setting comes from differences between two distinct technologies (rice and wheat), the results may be partly driven by differences in returns to the new technologies.[3] Furthermore, our model shows that heterogeneity can affect individuals' ability to learn from their neighbors, even when the distribution of characteristics causing the heterogeneity is known.

Finding that heterogeneity handicaps social learning also gives us additional confidence that the social network effects that we observe are due to learning rather than imitation. It is unlikely that we would observe a negative relationship between soil heterogeneity and social network effects if farmers were merely driven by conformity motives.

The paper is structured as follows: Section 2 situates the study in its context and makes the case that this environment is a suitable one in which to study social learning about a new technology. Section 3 presents a model of how heterogeneity in underlying characteristics affect individuals' ability to learn from others by increasing the variance of the posterior distribution

[2]Social-influence here denotes specifications where the main network variable is the number of treated in a farmer's network. Some studies refer to the peer-to-peer transmission of information about a technology as "nformation passing"; Banerjee et al., 2013 find that it is important in diffusing information about microfinance.

[3]Munshi notes that unlike early high-yielding rice varieties, the high-yielding wheat varieties had much higher returns than the traditional technology and had fairly certain yields.

of beliefs. Section 4 describes the data sources, presents summary statistics, and examines the exogeneity of the information randomization in the social networks. Section 5 presents details on the results and Section 6 concludes.

2 Context

The data for this paper come from a sub-sample of the study population of a large-scale randomized control trial (RCT) that aims to evaluate the socioeconomic impacts of a new hybrid maize seed that is produced by Western Seed Company (WSC). The main study centers on a cluster-randomized roll-out of information about and samples of the company's high-yielding maize hybrids.[4] Until recently, the company faced production capacity constraints and therefore had a limited geographic reach. The study villages are all located in what WSC considers to be expansion areas and that had neither had access to the seeds nor been exposed to information or marketing of the seeds. In other words, prior to the intervention most farmers in these areas had not heard of, let alone used, these hybrids. However, many of them will have some familiarity with hybrid maize seeds.[5] The intervention divided villages into treatment and control clusters; sampled farmers in treatment villages were invited to information sessions and given a 250g sample pack of WSC hybrid seeds.

Hybrid use among maize farmers in Kenya is quite widespread, unlike in many other parts of sub-Saharan Africa.[6] It may therefore seem like an unusual setting in which to study learning about hybrids as a new technology. Indeed, Suri (2011) uses the prevalence of hybrids in Kenya to motivate her assumption that farmers know the returns to hybrids, focusing instead on heterogeneity in farmers' returns to hybrid seed to explain differences in adoption. However, in doing so, she ignores the substantial diversity of seed varieties available in Kenya. For the numerous studies of technology adoption during the early days of the Green Revolution, it was reasonable to characterize a farmer's choice as being between a traditional technology and a high-yielding variety (HYV). In contemporary Kenya, however, where an average of over 14 new maize varieties have been released on the market each year since 2000, farmers face a much more complex choice.[7] The solid black line in Figure 1 shows the growing number of varieties being released on the Kenyan market since 1964, a number that ranges from 0 to 42 varieties per year (scale shown on right-hand y-axis). Further, the gray bars depict the reported yield ranges of the released varieties, illustrating the diversity of potential yields among the available seeds. This hints at a substantial – and growing – complexity

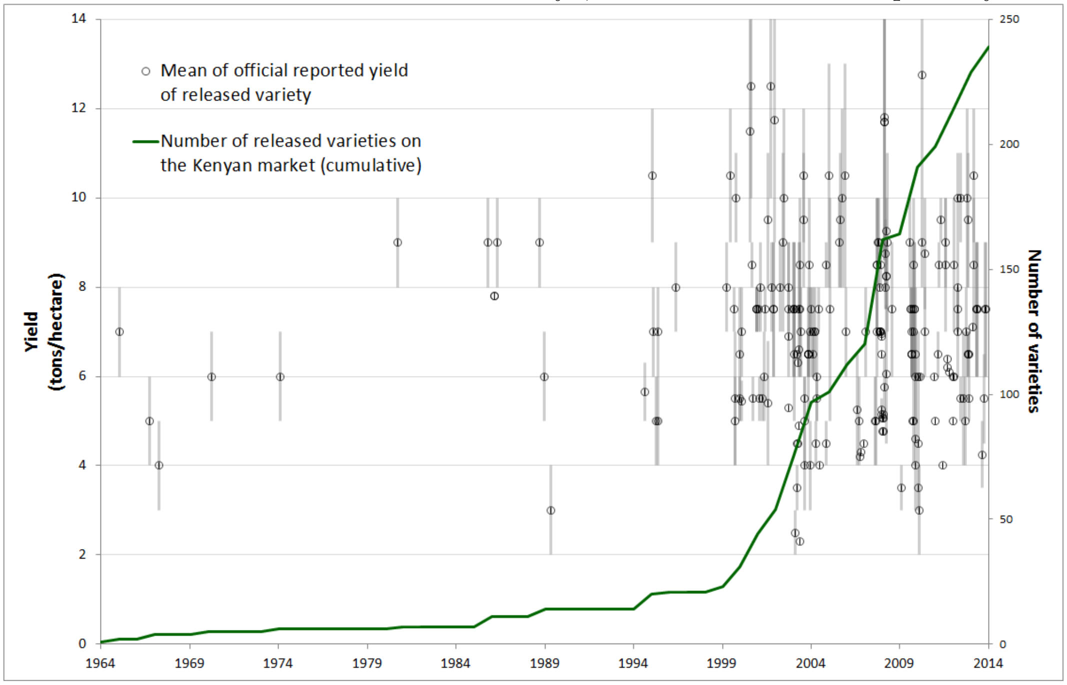
[4] We will not fully describe the RCT ("Evaluating the sociology-economic impacts of Western Seed's hybrid maize program") here, but more details on the study design can be found in Carter et al. (2014b).

[5] In the sample used for this paper, around 37% of the farmers had planted some form of hybrid in the four years prior to the treatment.

[6] Estimates range between 40 - 70%, depending on the region. In neighboring Uganda adoption rates are below 10%.

[7] Smale and Olwande (2014) report that there are currently 11 companies with maize varieties registered to their names, including local companies such as KSC and WSC as well as multinational and regional companies (for example Monsanto and Pannar, the latter based in South Africa). KEPHIS (2014) lists 240 hybrid varieties that have been released on the Kenyan market since 1964.

Figure 1: Number of maize varieties released in Kenya, 1964 - 2014 and their reported yield capacity



Source: Data from Kenya Plant Health Inspectorate Service (KEPHIS), 2014
Graph by author.

facing farmers when choosing which hybrid to plant.[8]

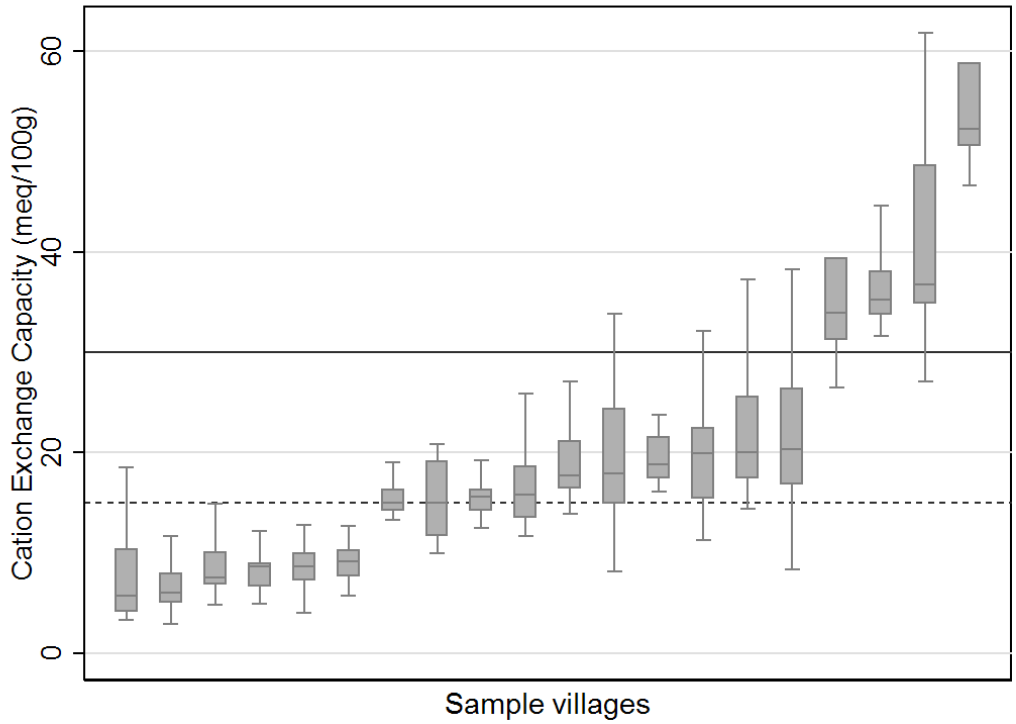
Despite this abundance of seed choice, many blame Kenya's stagnating maize production on how slowly older hybrids are being replaced by newer releases (see for example Mathenge et al. (2012)). In a nationally-representative panel survey collected in 2010, more than two-thirds of the farmers using hybrid maize seed planted a variety developed in 1986 (Suri, 2011). In the mid-1980s, the parastatal seed company, Kenya Seed Company (KSC) was the sole purveyor of maize hybrids in Kenya,[9] and Smale and Olwande (2014) note that these earlier hybrids were uniquely suited to the growing conditions of the Kenyan highlands. The authors also point out that these growing conditions are "unique by world standards" and, in particular, are quite distinct from the agro-ecological setting of our study area. In other words, these dominant older hybrids are not optimal for farmers in many parts of Kenya, leaving substantial room for improved yields if farmers adopt better-adapted varieties. We therefore believe that collapsing farmers' choices into "traditional" versus "hybrid" seeds neglects considerable complexity and ignores an aspect of learning that could be crucial in addressing the stagnating yields in sub-Saharan Africa.

Another important aspect of the setting of this study is that the region is characterized by significant heterogeneity in soil quality (Tittonell et al., 2008). This heterogeneity is driven both by inherent variation in soil types and by historical land use and management patterns. We collected

[8]A similar diversity can be seen in the optimal altitude ranges and months to maturity of the different varieties (shown in Appendix C).

[9]Their market share declined slowly between 1990-2000 and has since then dropped steadily to its current level of around 50%.

Figure 2: Box plot of CEC across the sample villages



soil samples on the main maize fields of each sample farmer in the RCT and had them analyzed by a leading agricultural soil testing lab. Figure 2 displays the variation in the Cation Exchange Capacity (CEC), a measure of the soil's ability to hold onto essential nutrients and to supply them to a crop (Jones and Jacobsen, 2005) that is often used to gauge a soil's fertility. The dashed and solid horizontal lines display the soil lab's definition of low and high CEC, respectively. As the figure shows, there is considerable variation both within and between villages. Importantly for our purposes, the villages also differ quite a bit in the *extent* of variation, which we exploit to study how heterogeneity affects learning.

3 Model

I model the yield-relevant characteristics of individual plots (such as soil type/variety, climactic conditions, elevation, rockiness of the terrain) as points on the positive real half-line. Each plot is represented by a real number and each farmer knows the type of her soil. Numbers that are close to each other on the real line represent plots that are similar in their characteristics. In this case, we can model the expected output of each plot, given a particular input such as fertilizer or hybrid seeds, as a realization of a Brownian motion process.

My modeling choice follows that of other authors who have modeled the space of possible technological innovations as a continuum. The payoff of each technology is a realization of a Brownian motion process, which guarantees that the productivity of similar technologies (i.e. technologies represented by points that are close to each other on the real line) is more closely correlated than that of more dissimilar technologies.

Jovanovic and Rob (1990) model technological progress by studying experimentation in the context of technologies modeled as realizations of Brownian motion. Since their seminal work, a growing number of papers model technologies in this manner. For example, Garfagnini and Strulovici (2016) model the payoffs to unknown technologies as realizations of a Brownian motion process over the real line. In the latter case, the authors are interested in contrasting "incremental" experimentation with radical experimentation to show that radical experimentation eventually stops and that technological progress would stagnate. In my setting, rather than allowing agents to choose where along the real line to experiment, and rather than having infinitely many technologies, I model agents as experimenting with one technology that has different returns on different types of plots, represented by a pre-determined soil quality, c.

Assume thus that farmers have incomplete knowledge of the returns to the new technology and that payoffs from the technology are some (unknown) function $f(c)$ of the underlying environment c. In our setting, the environment in question is soil CEC, and farmers can observe others' experiments with the technology to learn about the shape of $f(\cdot)$. Farmer i in village g has a fixed soil quality c_i, which is drawn from the positive real half-line, with variance σ_c^2, and the soil qualities of her neighbors in the village are all drawn from the same distribution. Assume further that the unknown payoffs are realizations of a one-dimensional Brownian motion with drift $\kappa = 0$ (i.e. no-drift) and normalized variance $\sigma = 1$. As is standard, I assume $f(0) = 0$.

Farmer i knows her own quality, c_i as well as the qualities of others in her village. She also observes a sample of signals from n of her village neighbors. Farmer i's neighbors' soil qualities are thus the sample $s = \{c_{i1}, c_{i2}, ..., c_{in}\}$, and the yield outcomes that i observes are $y = \{f(c_{i1}), f(c_{i2}), ..., f(c_{in})\}$.

Proposition 1. *The variance of farmer i's posterior belief about $f(c_i)$ is increasing in the variance of soil quality in the village, σ_c^2.*

The proof of Proposition 1 is in Appendix B, but the intuition is straightforward: In homogeneous villages, i.e. where soil quality varies little, the nearest-neighbor of farmer i is more similar to her (i.e. closer to her along the dimension of soil quality) than in a more heterogeneous village. Farmers in homogeneous villages can therefore update their beliefs about the shape of their own production function, $f(c_i)$, with more precision than can farmers in heterogeneous villages. Unlike Munshi (2004), this model does not require the farm level characteristics to be unobservable for heterogeneity to slow learning.

4 Data

The main randomization of information and seed samples was conducted at the cluster-level (clusters of 3 villages), but the relevant treatment for this study was actually randomized at the household level. Households were randomly selected in both treatment and control villages for inclusion in the study.[10] In the treatment villages, the sample households were invited to an information session and given a 250g sample pack of the new seeds. While the main goal was to induce differences in adoption levels between treatment and control villages, the design also resulted in variation *within* treatment villages in the level of experience with the new technology – variation that should be orthogonal to farmer attributes, including their social network characteristics.

The rest of this section describes the data and defines the various variables used in the analysis. First, we describe who our sample farmers are, their relationship with the larger RCT sample, and the various data sources used in the study. We then provide some additional information on the construction of our measure of the treated farmers' experiences with the technology. Finally, we show summary statistics and balance tests, comparing means across the treated and indirectly treated as well as showing that the number of treated links in an individual's network are as good as random (with some notable exceptions that we subsequently control for in our analysis).

4.1 Data sources

The RCT described above conducted an extensive baseline survey with the sample farmers in both treatment and control villages, collecting data on a variety of demographic and agricultural indicators as well as soil samples. For this paper, a second pool of farmers was included – a random selection of *untreated* farmers residing in treatment villages. We refer to these farmers as the indirectly treated. Table 1 shows what components of the study the different samples participated in. No network data was collected in control villages and the indirectly treated were not included in the baseline survey. Henceforth, we will only discuss data on the directly and indirectly treated farmers unless otherwise noted.

Figure 3 shows the timeline of the various data sources used for this study. The information sessions, during which the sample packs of seed were distributed, were held before the main planting season of 2013.[11] Treated farmers therefore had the opportunity to plant the sample seeds (and to learn from their performance) during the main season and more than 97% of the farmers report having planted the samples. The baseline survey was then conducted in treatment and control villages in November 2013. We did not expect to see impacts on any of the key indicators (yields, incomes, and food security) by the time of the baseline survey as the sample packs only contained enough seeds to plant a very small experimental plot (about $\frac{1}{30}^{th}$ of the average farmer's land). The lack of impact at the time of the baseline was confirmed by t-tests of the equality of the main indicator variables between treated and control households (Tegemeo, 2014).

[10] The author conducted the within-village randomization based on listings of the full set of households in the village, from which approximately 15 were drawn without replacement using an Excel spreadsheet. The exact number of households sampled by village was decided to ensure that the sampling probability was proportional to the size of the village, within the cluster. Clusters are made up of 3 villages.

[11] Kenya has a bimodal rainfall distribution, with most of the country (and the areas in this study) experiencing long rains in March - May and short rains October - December. The long rains are called the main season, and the short rains are called the short season.

Table 1: Farmer types

Farmer status	Village	Info + sample	Baseline	Soil sample	Network
Directly treated	Treatment	Yes	Yes	Yes	Yes
Indirectly treated	Treatment				Yes
Control	Control		Yes	Yes	

4.1.1 Phone survey

In early 2014, we conducted a phone survey with the treated farmers, eliciting both objective and subjective information about their experiences with the sample seeds as well as asking respondents whether they knew where to purchase the product if they wanted to buy the hybrids for the coming main season. We control for the proportion of respondents in each village who know where to purchase the seeds in all our regressions, but we would expect it to have the greatest impact on hybrid purchasing behavior.

Based on respondents' answers to two questions about the performance of the sample, we construct a measure of her experience with the seeds that we refer to as the "perceived experimental gain". In particular, we are interested in two questions from the phone survey: "How much did you harvest from the sample pack seeds?" and "How much maize would you have harvested (assuming the same weather, input use, etc.) if you had planted the seeds you normally grow instead of WSC hybrids?" We use the percentage increase over a respondent's "expected" harvest to measure the result of her on-farm experiment with the seeds; more details are provided in Section 4.2.

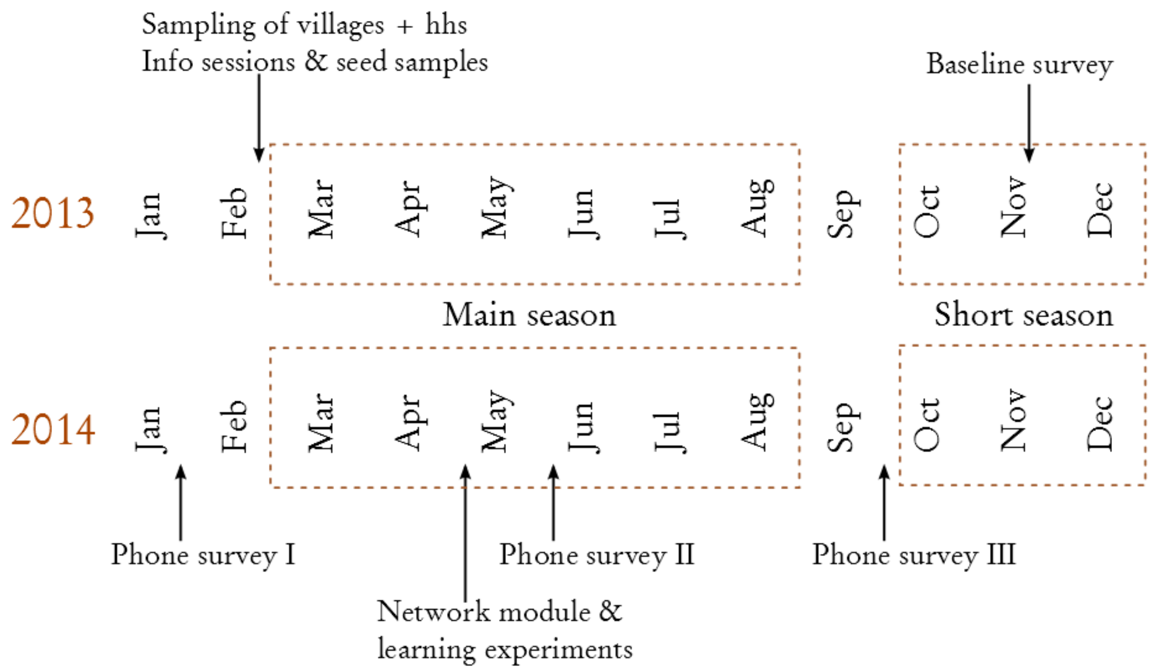
4.1.2 Network module

The data on social networks come from an add-on survey conducted in a sub-sample of 20 treatment villages in Western Kenya, conducted in May 2014, after the planting of the main season. All the treated farmers in a village were invited to an experimental session together with a random sample of their fellow (untreated) villagers. The total number of farmers invited to each session was 30, so since there is slight variation in the number of treated farmers in each village, there is also some variation in the number of indirectly treated farmers.[12] This should not affect the results as all regressions control for total network size.

Out of the 600 farmers invited to the sessions, 575 (96%) showed up and participated. Together, these 575 farmers make up the experimental sample. Each farmer answered a short entry survey when she arrived, eliciting information on input use in the current season (including what maize varieties she planted). Since the indirectly treated farmers were not included in the baseline survey, they were asked some additional questions during the entry survey (including land size, history of

[12] This was done for the purposes of an experimental game that the invited farmers participated in, and which required the total number of farmers present to be as constant as possible across sessions. Overall, the total number of respondents per village ranges from 26 to 30, and the share of treated farmers ranges from 26% - 84%, with a mean of 56%.

Figure 3: Timeline of data collection



fertilizer and hybrid use, and the questions used in computing the Progress out of Poverty Index). These questions were largely the same as the ones used in the baseline survey, with some notable exceptions that are discussed in Section 4.3.

The farmers were presented with a network module, in which they answered a variety of questions about the other 30 farmers present. The questions include the identity of any relatives, who they speak to about agricultural issues (at different frequencies), geographical proximity, and perceived similarities. Another set of questions elicit their knowledge about other farmers in the experimental sample (e.g., what maize seed they planted and whether they would recommend WSC hybrids).[13] Throughout the analysis, individual j is counted as being in person i's social network if person i listed her in *any* of the network questions.[14]

As part of the survey instrument during the experimental sessions, we tried a novel method of eliciting farmers' willingness to pay (WTP) for WSC seeds that we call a price-premium approach.[15]

[13]The full list of network questions can be found in Appendix A.

[14]The network module was conducted using tablets. During the entry survey, all farmers were photographed and the close-ups were presented in a touch-enabled matrix for respondents to simply select their answers to each question. This provided a highly intuitive and relatively quick way of collecting network data, which enabled us to ask as many network questions as we did.

[15]Since the seeds are available on the market, an incentive-compatible experimental auction was infeasible. We pre-tested slightly simpler hypothetical WTP-questions, constructing hypothetical scenarios such as "If you had to send a friend to the market to buy you WSC hybrids but you were unsure how much they cost, what is the maximum amount of money that you would send with him/tell him to pay for a 2-kg bag?", etc. These types of questions were very difficult to communicate to farmers and all the respondents either objected that they do not send others to buy seeds for them or were very confused as to why they would want to pay more than the market price. Responses such as "But I know that the price is x, why would I tell him to pay more?" were very common.

The first step entailed asking respondents to identify the seeds with which they are familiar. To keep the list of seed choices manageable, they could choose from a list of 17 seeds, which included all the seed varieties that more than 2 farmers in the sample villages had used in the previous year (based on the baseline data). Among the seeds were 4 Western Seed varieties: WS 505, WS 507, WS 509 and WS 303. Familiarity was defined more narrowly than simply having heard of the seeds: a respondent was defined as being familiar with the seed if *(i)* she, or someone she knows, had used it or *(ii)* she had heard about the seed *and* felt that she knew how it compares to other seeds.

Second, having identified the seeds with which each respondent was familiar, we asked her to rank these seeds. The enumerator displayed laminated cards with the names of the seeds that the respondent expressed familiarity with and asked "If all these seeds cost the same amount, which would be your first choice?". The laminated card representing each farmer's first choice was then removed, and she was asked about her second choice, and so on until she had ranked all of the varieties with which she was familiar.

Third, for all farmers who ranked at least one WSC variety above a non-WSC variety, we elicited the premium that they would be willing to pay over the non-WSC variety.[16] For respondents who found this question difficult to answer, we asked them to imagine that they held a 2-kg bag of the non-WSC variety and that the enumerator had a 2-kg bag of the WSC variety, and then asked how much they would pay the enumerator to trade bags. By adding the premium to the price of the non-WSC variety used for comparison, we obtain the respondent's WTP for the WSC variety.[17]

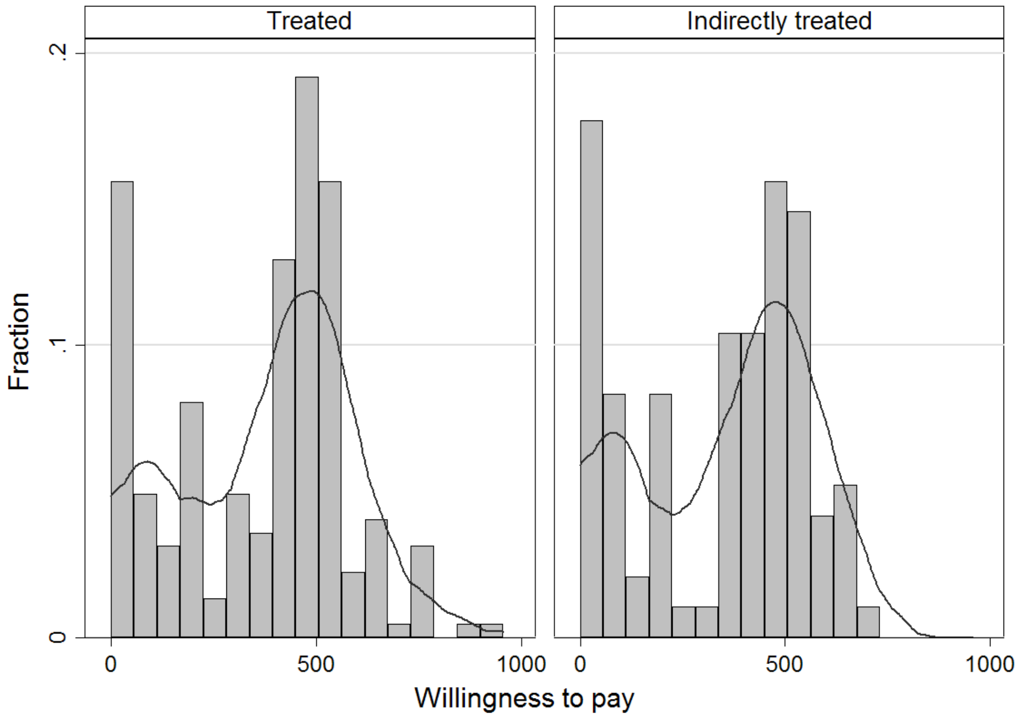
This price-premium strategy limits the number of farmers for whom we observe WTP (we only obtain a ranking for farmers who are familiar with the seeds, and even among those with a non-missing ranking, it was possible to rank a WSC variety below farm-saved seeds, resulting in a non-positive premium as the cost of these saved seeds is essentially 0). We believe, however, that this was the only defensible way to elicit WTP in this context and the distributions and results for WTP give us confidence that this variable is indeed providing a useful measure of farmers' true willingness to pay. Figure 4 shows the distribution of WTP by treatment status. The distributions are quite similar – overall it seems that, conditional on being familiar with the seed, the indirectly treated do not differ greatly in their WTP.[18]

[16] If they mentioned more than one WSC variety, we asked about the premium compared to the highest-ranking WSC variety.

[17] This methodology appeared to do well in getting around the issue of anchoring on the market price; farmers did not seem confused about why we were asking them this. Further, at least anecdotally, farmers don't seem to be particularly subject to hypothetical bias: several respondents came up to the survey team at the end of the surveys and asked if they could now pay the premium and trade in for a bag of WSC. Finally, as an informal check of the WTP-WTA gap, a small sub-sample of farmers were asked (after their premium had been recorded) what their answer would be if instead we turned the question around (i.e. they had the 2-kg bag of WSC maize and the enumerator had a 2-kg of the non-WSC variety – how much would the enumerator have to pay them to trade?). They consistently looked confused and exclaimed "The same!"

[18] A Kolmogorov-Smirnov test for equality of the two distributions fails to reject the null of equality ($p-value:\ 0.52$)

Figure 4: Histogram of willingness to pay for Western Seed varieties, by treatment status



4.2 Experience with the technology

Our measure of the treated farmers' experiences with the sample pack seeds is computed using two questions from the phone survey.[19] We define farmer i's experience with the technology as follows:

- Farmer i's actual experience (y_i) is based on the survey question "How much did you harvest from the sample pack seeds?"

- Farmer $i's$ subjective counterfactual ($\tilde{y}_i$) is based on her answer to "How much would you have harvested (given the same weather, input use, etc.) if you had planted the seeds that you normally grow instead of WSC hybrids?"

We then calculate i's perceived experimental gains, Δ_i as

$$\Delta_i = \frac{y_i - \tilde{y}_i}{\tilde{y}_i} \tag{1}$$

Our measures of the distribution of signals flowing through a person's network consist of the mean and variance of the perceived experimental gains, i.e.:

[19]More details on the phone survey are in Section 4.1. Out of the 320 treated participants in the network module, 292 (91%) were reached by the phone survey. The 28 farmers that the enumerators failed to reach do not appear to be any different than the ones who answered the phone survey, based on t-tests of baseline characteristics. We have also re-estimated our results with a data set in which the missing farmers' experiences were replaced by the mean experience in their village. The results using this 'imputed' sample are consistent with those obtained using the raw data, and some network effects appear stronger when experiences are imputed.

$$
\begin{aligned}
\mu_i &= \sum_{j \in N_i} \frac{\Delta_j}{N_i} \\
\sigma_i &= \sum_{j \in N_i} \frac{(\Delta_j - \mu_i)^2}{N_i}
\end{aligned} \tag{2}
$$

where $j \in N_i$ denotes that farmer j is in farmer i's network.[20]

Figure 5 shows the overall distribution of Δ_i for the treated farmers as well as the μ_i for both the directly and indirectly treated.[21] The distributions of average signals resemble that of the overall distribution, suggesting that even if networks are endogenous (discussed in more detail in Section 4.4), farmers did not select with whom they spoke based on whether or not they had success with the sample seeds. In fact, if anything, the average signal in individuals' networks is somewhat more compressed around low average signals than the overall distribution.

4.3 Summary statistics

Summary statistics on the experimental sample are presented in Table 2. The table also presents a t-test of equality of means of these variables between the treated and the indirectly treated. These t-tests are included for completeness only, since whether or not the treated and the indirectly treated respondents are similar on observables is not the key concern for the social network analysis, especially since we run all the network analyses separately for the treated and indirectly treated.

Risk preferences were measured using a question asking people about their willingness to take risks "in general,"[22] following Dohmen et al. (2011), who validate this type of survey question against an incentivized lottery and find that the general risk question is a reliable predictor of actual risk-taking behavior and risk attitudes. The understanding score comes from a series of questions designed to test respondents' understanding of the experimental games and is constructed as the share of questions that each respondent answered correctly. The questions were of varying difficulty, from a very basic understanding of what the game represented to more challenging questions about how to read an empirical probability distribution function.

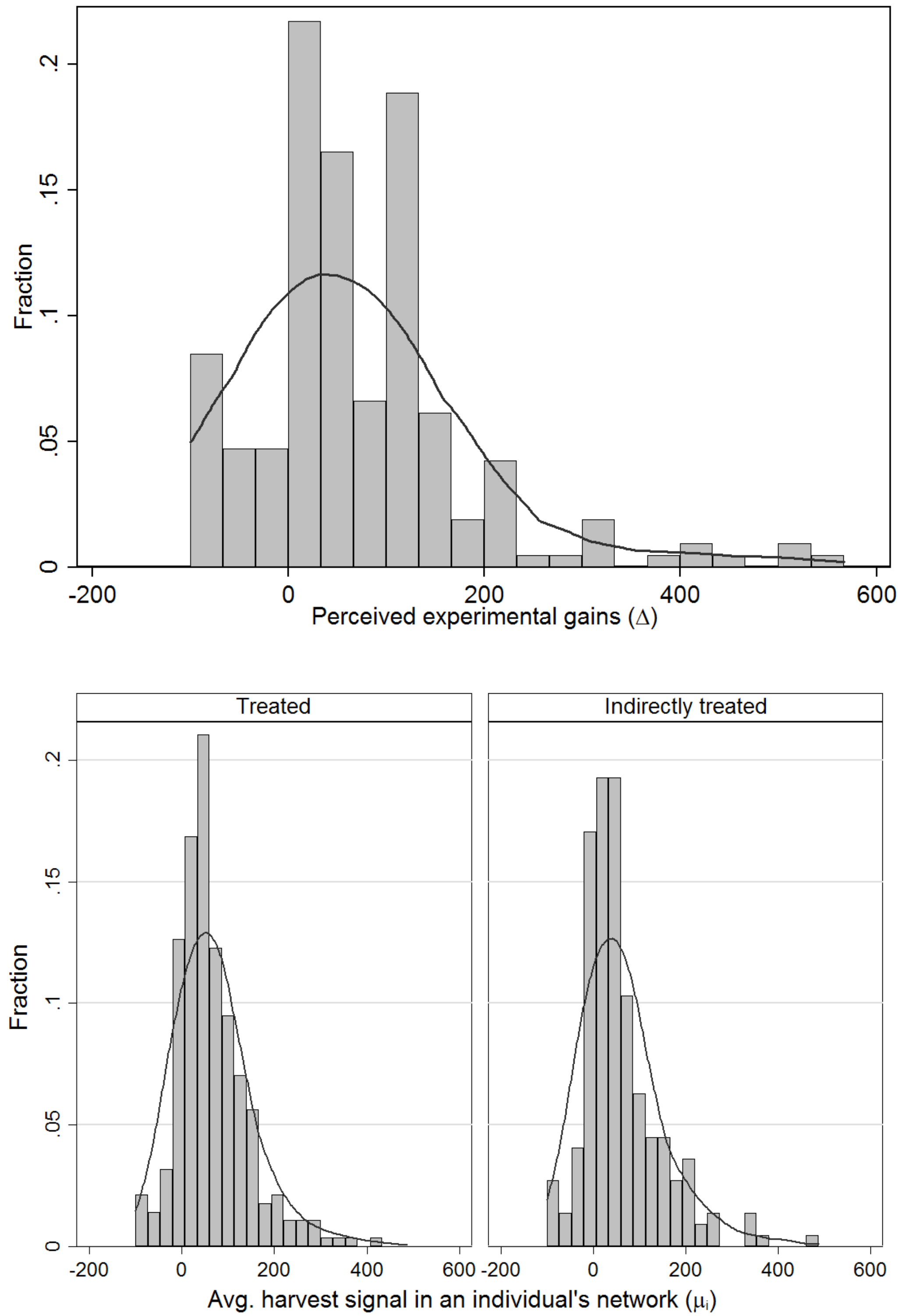
The Progress out of Poverty questions are taken directly from the Kenyan scorecard designed to compute the PPI index. The index is based on 10 questions that have been found to be significant predictors of socioeconomic status in a nationally representative household survey. Households answer 10 basic questions that get converted into a score between 0-100 and that can be used to find the household's likelihood of being below various poverty lines. A higher score on any given

[20]The denominator for σ_i is N rather than $N-1$ since σ_i is the population standard deviation – I observe Δ for the population of treated farmers in their network, not just a sample.

[21]The histogram excludes a few outliers for clarity. A t-test fails to reject that the means of the distributions for treated and indirectly treated differ, with a p-value of 0.56.

[22]The wording of the question was as follows: "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?", where 0 was labeled as "Not at all willing to take risks", and 10 meant "Very willing to take risks".

Figure 5: Distributions of treated farmers evaluation of the performance of the hybrid seed samples
The perceived experimental gains (Δ) have been multipled by 100, so these are in percent



question indicates a lower probability of being below a poverty line. Overall, the mean score in the sample is just below 45, which according to the PPI methodology translates into the average farmer having a roughly 15% probability of being below USAID's "extreme" poverty line, a 37% probability of being below Kenya's national poverty line, and an 83% probability of being below a $2.50/day poverty line.[23]

In general, the treated and indirectly treated look very similar on the above variables, giving us confidence in the random sampling within villages. However, as Table 2 shows, some of the agricultural characteristics differ between the two groups. In particular, the size of households' main maize fields and the number of seasons that they used fertilizer in the four years (8 seasons) preceding the treatment are significantly different. The treated households appear to have maize fields that are about 0.2 acres (0.08 hectares) larger and have used fertilizer about half a season more (out of 8) than the indirectly treated. These differences are almost certainly artifacts of the different data sources used. While the PPI and other household characteristics were collected using identical protocols, the agricultural information for the treated households was collected using an extremely detailed agricultural module in the baseline survey.[24]

As for network characteristics, the treated and indirectly treated have about the same number of relatives (2.4 relatives, with 1.3 of them being treated). This network measure is likely the most fixed among the social networks elicited and is that which comes closest to a measure of the baseline social network. As explained in Section 4.1, the social network module took place more than a year after the information sessions were held and the sample packs distributed. It is therefore quite possible that the social networks in these villages reacted to the treatment. This does seem to be the case judging by the other network characteristics in Table 2.

The number of links refers to the sum of links across all the social networks investigated. Individual j is in individual i's network if i mentions j, while the reciprocal version only counts those links in which i mentions j *and* j mentions i. On average, sample farmers have links to 7 other farmers (out of the 30 in the sessions), 4 of which are treated. This number drops by half when we only consider reciprocal/strong links, a pattern that also holds for the number of treated links. It appears as though networks responded substantially to the treatment since the treated have more links to other treated households.

In an attempt to control for this issue, we constructed a "corrected" network variable. We asked respondents which of their contacts they had spoken to about maize farming *for the first time* in the months since the treatment and then removed those new links from the main network. By so doing, we anticipated picking up potential endogenous network formation and obtain an ex ante network variable. As the table shows, this strategy only removed less than a third of a link – about 4.5% (5.6%) of the total number of (treated) links. That the treatment affected networks

[23]Please see http://www.progressoutofpoverty.org/ for detailed information on the methodology and examples of scorecards for different countries.

[24]Land size, for example, was collected based on a listing of all parcels that the household cultivates, followed by each field on each plot (where field is essentially a section of the parcel, defined by a uniform crop – or set of crops, in the case of inter-cropping), and then the sizes of each field were elicited. For the indirectly treated, this information was gathered based on a question in the entry survey eliciting, simply, the size of their main maize field.

Table 2: Summary statistics & t-test for differences in means between treated and indirectly treated

Variable	mean	sd	min	max	mean(T) - mean(I)	t-stat
Household characteristics						
Kiswahili spoken at home	0.03	0.18	0	1	-0.001	(-0.06)
Luhya spoken at home	0.19	0.39	0	1	0.045	(1.41)
Luo spoken at home	0.78	0.42	0	1	-0.045	(-1.29)
In womens' or farm group	0.48	0.50	0	1	0.076*	(1.83)
In microfinance group	0.25	0.43	0	1	0.009	(0.25)
General risk taking attitude (0-10)	8.15	2.04	0	10	0.081	(0.47)
Understanding score, exp. games	0.74	0.34	0	1	-0.024	(-0.85)
Rooftop, main dwelling (grass/iron sheet)	0.85	0.36	0	1	0.030	(1.01)
Progress out of Poverty Index						
Sum of core 10 PPI scores (0-100)	44.49	12.41	14	84	1.409	(1.35)
No. of hh members (0-32)	10.34	8.60	0	32	0.940	(1.31)
Educ. of female head / spouse (0-11)	3.36	3.09	0	11	0.386	(1.51)
Occupation of male head / spouse (0-9)	6.37	2.76	0	9	-0.341	(-1.50)
No. rooms in main dwelling (0-8)	3.84	2.37	0	8	0.097	(0.49)
Floor of main dwelling (0-3)	0.76	1.31	0	3	0.058	(0.53)
Main source of lighting fuel (0-12)	7.05	2.39	0	12	-0.009	(-0.04)
Owns irons (charcoal or electric)? (0-4)	2.15	2.00	0	4	0.086	(0.51)
No. of mosquito nets owned (0-4)	3.45	1.13	0	4	0.035	(0.37)
No. of towels owned (0-10)	4.85	4.43	0	10	0.271	(0.73)
No. of frying pans owned (0-7)	2.33	2.04	0	7	-0.113	(-0.66)
Agricultural characteristics						
Size of main maize field (acres)	1.30	1.16	.07	10	0.201**	(2.16)
No. of seasons used fertilizer, 4 years	2.57	3.33	0	8	0.479*	(1.71)
No. of seasons used hybrids, 4 years	3.32	3.33	0	8	-0.059	(-0.21)
Network characteristics						
No. of relatives	2.43	2.23	0	12	0.070	(0.38)
No. of treated relatives	1.31	1.39	0	8	0.080	(0.69)
No. of links (all)	7.05	3.92	0	29	0.344	(1.08)
No. of treated links (all)	4.08	2.51	0	20	0.549***	(2.69)
No. of reciprocal links (all)	3.29	2.50	0	22	0.409**	(2.01)
No. of treated reciprocal links (all)	1.93	1.71	0	15	0.435***	(3.15)
No. of links in corrected network	6.73	3.78	0	29	0.154	(0.50)
No. of treated links, corrected network	3.85	2.41	0	19	0.400**	(2.03)

The PPI variables are converted into a score; higher numbers indicate a lower probability of being below a poverty line.
t statistics in parentheses, standard errors clustered at the village level
* p<.1, ** p<.05, *** p<.01

is not surprising and the difference in number of links between the two groups should not affect the results since we run our analysis separately for the treated and indirectly treated. As the next section shows, the number of treated links seems fairly well randomized across baseline variables. Therefore, even if networks are endogenous we might worry less about correlated unobservables.

4.4 Balance on observables

The key aspect of the randomization for this type of network analysis is not whether the treated and indirectly treated are similar. Instead, our approach requires that the treatment-induced variation in the number of treated network members that a given individual has is exogenous. In other words, we require that, conditional on individual i's total number of links (total network size), the number of *treated* links was randomized. We test the validity of this assumption by regressing baseline characteristics on the number of treated links, controlling for the total network size. Table 3 shows the coefficients on the number of treated links.

We undertake this test separately for the treated (T) and indirectly treated (I) for three different network variables: relatives (columns 1 & 2), the full uncorrected network (columns 3 & 4), and the corrected network (columns 5 & 6). The first two columns constitute the best test of whether or not the randomization was successful as we can be fairly certain that any systematic significant differences are due either to faulty randomization or to chance since the number of relatives that our sample farmers have is clearly independent of the treatment.

Most of the baseline variables do not differ significantly by the number of treated relatives for each respondent. The few significant variables are in line with what we would expect to occur by chance and they are not very large in magnitude. The one exception is that *indirectly* treated farmers who have more treated relatives are significantly more likely to have used hybrids. It is worth noting that having one additional treated relative represents close to a doubling of the number of treated relatives compared to the sample mean of 1.31. Hybrid use and the number of seasons a household had used fertilizer is also significantly larger for households with more treated links in the other two networks. For the other social networks, this difference might be explained by respondents who had used improved inputs in the past seeking out the treated when they heard that they had received information about a new technology, but it is harder to explain for relatives. Bandiera and Rasul (2006) similarly observe that adopters have significantly more adopters among their friends and family, in a non-random sample of villages exposed to a new agricultural technology in northern Mozambique. The authors do not explore this fact in more detail.

For the other two network measures, the randomization seems similarly credible. The few differing PPI variables, while significant, are tiny in magnitude as the PPI is scored out of 100. In addition, the overall PPI score is not significantly different by treated links. We address the significant differences by including fertilizer and hybrid history as control variables in all our regressions.

Table 3: Balance test: Regression of baseline vars on number of treated friends

Variable	Relatives T	Relatives I	Full T	Full I	Corrected T	Corrected I
	Coeff. on no. of treated links, controlling for nr. of links					
Household characteristics						
Kiswahili spoken at home	-0.009	0.006	0.003	0.013	0.004	0.013
	(-1.54)	(1.02)	(0.59)	(1.40)	(0.73)	(1.34)
Luhya spoken at home	0.086*	0.011	0.027	0.031	0.021	0.027
	(1.84)	(0.62)	(1.02)	(1.53)	(0.84)	(1.41)
Luo spoken at home	-0.077	-0.017	-0.030	-0.045*	-0.024	-0.041
	(-1.64)	(-0.94)	(-1.07)	(-1.71)	(-0.92)	(-1.59)
In womens' or farm group	-0.009	0.008	-0.006	0.021	-0.012	0.024
	(-0.20)	(0.23)	(-0.32)	(1.02)	(-0.59)	(1.21)
In microfinance group	-0.047*	-0.002	-0.005	0.041***	-0.013	0.040***
	(-1.90)	(-0.07)	(-0.38)	(3.60)	(-0.89)	(3.57)
General risk taking perception (0-10)	-0.089	0.018	-0.039	-0.039	-0.061	-0.033
	(-0.50)	(0.12)	(-0.62)	(-0.41)	(-1.03)	(-0.34)
Understanding score, exp. games	-0.010	0.035	-0.007	0.018	-0.012	0.017
	(-0.42)	(1.33)	(-0.71)	(0.88)	(-1.16)	(0.88)
Rooftop, main dwelling (grass/iron sheet)	0.024*	0.010	0.014	0.025	0.007	0.025
	(1.67)	(0.40)	(1.15)	(1.45)	(0.53)	(1.42)
Progress out of Poverty Index						
Sum of core 10 PPI scores (0-100)	-0.506	1.248	-0.317	0.835	-0.354	0.655
	(-0.68)	(1.09)	(-0.50)	(1.27)	(-0.52)	(1.02)
No. of hh members (0-32)	-0.587	-0.413	-0.616**	-0.489	-0.673**	-0.560
	(-0.79)	(-0.71)	(-1.98)	(-1.16)	(-1.97)	(-1.27)
Educ. of female head / spouse (0-11)	-0.035	0.369	0.204	0.308**	0.219	0.257*
	(-0.11)	(1.61)	(1.17)	(2.15)	(1.32)	(1.83)
Occupation of male head / spouse (0-9)	0.165	0.231	0.008	0.055	0.074	0.034
	(0.78)	(0.90)	(0.09)	(0.34)	(0.82)	(0.22)
No. rooms in main dwelling (0-8)	0.052	0.236	0.045	0.180	0.027	0.183
	(0.30)	(1.42)	(0.37)	(1.35)	(0.21)	(1.40)
Floor of main dwelling (0-3)	-0.026	0.144	0.041	0.080	0.032	0.080
	(-0.37)	(1.40)	(0.51)	(1.09)	(0.36)	(1.08)
Main source of lighting fuel (0-12)	0.037	0.209	0.029	0.069	0.012	0.072
	(0.20)	(1.07)	(0.20)	(0.54)	(0.08)	(0.56)
Owns irons (charcoal or electric)? (0-4)	-0.110	-0.223	-0.057	0.075	-0.059	0.063
	(-0.85)	(-1.07)	(-0.54)	(0.57)	(-0.55)	(0.47)
No. of mosquito nets owned (0-4)	0.214**	0.002	0.112**	0.046	0.107**	0.049
	(2.13)	(0.02)	(2.55)	(0.68)	(2.27)	(0.70)
No. of towels owned (0-10)	-0.214	0.484	-0.117	0.311	-0.107	0.276
	(-0.70)	(1.00)	(-0.54)	(0.97)	(-0.48)	(0.85)
No. of frying pans owned (0-7)	-0.002	0.210	0.034	0.200**	0.014	0.202***
	(-0.01)	(1.60)	(0.37)	(2.46)	(0.14)	(2.58)
Agricultural characteristics						
Size of main maize field (acres)	-0.026	0.024	-0.022	-0.042	-0.029	-0.038
	(-0.27)	(0.35)	(-0.42)	(-0.75)	(-0.55)	(-0.69)
No. of seasons used fertilizer, 4 years	0.440	0.271	0.312*	0.554***	0.303	0.536***
	(1.37)	(1.07)	(1.69)	(3.19)	(1.56)	(3.21)
No. of seasons used hybrids, 4 years	0.334	0.882***	0.295	0.664***	0.244	0.628***
	(1.26)	(2.92)	(1.63)	(4.15)	(1.32)	(3.88)

t statistics in parentheses, standard errors clustered at the village level

* p<.1, ** p<.05, *** p<.01

5 Empirical results

5.1 Empirical strategy

We will conduct the bulk of our analysis of social network effects using four main outcome variables:

1. *Familiarity with WSC hybrids*, an indicator variable equal to 1 if a respondent is familiar with the technology. The variable comes from the first stage of the WTP module (described in detail in Section 4.1), where respondents selected the seeds about which they felt they had sufficient knowledge to compare to other seeds. This outcome variable is closely related to the concept defined as "information passing" in the study by Banerjee et al. (2013), which focuses on the diffusion of microfinance through networks.[25] Intuitively, an individual needs to be familiar with the new technology before she can adopt it.

2. *WTP for WSC hybrids*, a respondent's stated willingness to pay for the technology, as described in Section 4.1. While stated WTP-measures have limitations, we believe that this measure could pick up learning if adoption impacts are limited by liquidity constraints and/or market imperfections. Such constraints might prevent farmers from purchasing a technology even when they have learned that it would be profitable for them. Due to the methodology used, we only have the WTP outcome for a subset of the sample.

3. *Planted a WSC variety*, a dummy variable for whether the household planted a WSC hybrid during the main season of 2014. This is our main measure of adoption. It is worth noting that it is a significantly more stringent measure of adoption than many used in the recent experimental literature on network effects and technology adoption. In Bandiera and Rasul (2006), Cai et al. (2015), Oster and Thornton (2012), and Miguel and Kremer (2004), the product under study was available to all study participants for free or at a heavy subsidy,[26] so that variation in adoption in these cases is more likely to stem from a desire or ability to use the product rather than from an assessment that it is profitable. In other words, the participants in these studies did not face many of the real-world constraints to adoption (for example, the need to finance the initial purchase), making the cost of adoption much lower than in our setting.

4. *Planted a non-WSC hybrid*, a dummy for whether the household planted a hybrid from a different seed company in the main season of 2014. If farmers learn from their treated peers that WSC hybrids are more profitable than other hybrids, then we could either observe no network effects on the probability of planting other hybrids (especially if many farmers were previously planting non-hybrids) or a negative impact (if farmers are substituting away from

[25] While the authors study the effects of the *network position* of the initial recipients of information, they are also trying to better understand how information about a new product diffuses through a network and what factors influence whether individuals in the network adopt the product.

[26] The weather insurance product in Cai et al. (2015) was heavily subsidized: the post-subsidy price only amounted to around 1% of production costs.

other hybrid varieties towards the WSC hybrids). There could also be positive spillovers if farmers infer from a positive signal about WSC hybrids that other hybrids are also worth trying, especially if WSC hybrids are unavailable in some communities.

5.1.1 Social networks

In its most general form, our econometric specification is the following:

$$y_{iv} = f(N_{iv}) + \gamma \mathbf{X}_i + \varepsilon_{iv} \tag{3}$$

where y_{iv} is one of our four outcome variables for household i in village v, $\mathbf{X}_i$ is a vector of baseline control variables, which are described below, and $f(N_{iv})$ is some function of the information in individual i's network.[27] We expect outcomes to be correlated within villages, and therefore cluster all standard errors at the village level.

We examine two main measures of information in person i's network: N_{iv} represents either *(a)* the number of treated farmers in her network (the social influence model), or *(b)* first two moments of the distribution of experiences reported by the treated individuals in i's network (called signal regressions). As described in Section 4.2, we measure the performance in how much higher (in percentage terms) the harvest from the sample pack was, relative to the expected harvest had she planted her regular seeds. The assumption is that the experiences of the farmers in person i's network combine to form a distribution of signals from which she can learn.

There are numerous possibilities for the form of $f(\cdot)$, and little consensus in the literature. For regressions akin to the ones we call social-influence regressions (i.e. using the number of treated links), different studies have assumed different functional forms. On one extreme, Babcock and Hartman (2010) and Oster and Thornton (2012) assume linearity in the effects of treated peers by simply including a variable for the number of treated people in person i's network.[28] Cai et al. (2015) impose a different assumption by using the share of treated in the network, while Carter et al. (2014a) instead opt for a very flexible functional form, including multiple indicator variables (for respondents having one, two, three, four, or five or more treated farmers in their network, and the same number of indicator variables for social network size). Magnan et al. (2013) have very sparse networks (the average farmer identified less than one agricultural contact in his village) and therefore use a dummy variable for the presence of at least one adopter in a person's network.

We choose an approach in-between the latter two papers in the social influence regressions, including indicator variables for one, two or "three or more" treated network members.[29] For the regressions that examine the signals that farmers receive, we include the first two moments of the

[27] Blume et al. (2011) and Breza (Forthcoming) note, respectively, that theory provides little guidance as to the correct measure of the groups that matter in social interactions, and that there is no consensus definition of what constitutes a link between two nodes in a social network. We therefore use the most conservative definition of a social network for our analysis, i.e. the (corrected) 'uber-network'. In other words, individual j is counted as being in person i's social network if person i identified her as being in *any* of her networks.

[28] Oster and Thornton (2012) also conduct robustness checks that include the share of treated peers in the network.

[29] The results are similar when we include more dummies and suggest decreasing marginal effects of additional treated links.

distribution of signals in a respondent's network.[30] Using the average signal imposes symmetry between reports of negative and positive experiences; we have also tried specifications allowing positive and negative signals to have different weights, which does not substantially change the results.[31] We expect that the average signal in an individual's network will make them more likely to adopt the new technology. On the other hand, more variability will decrease the precision of the signal and make the individual less likely to react to it.

More specifically, then, we estimate the following two basic models:

$$y_{iv} = \alpha_1 + \beta_k \sum_{k=1}^{K} l_{iv}^k + \gamma_1 \mathbf{X}_i + \varepsilon_{iv} \tag{4}$$

where the l_{iv}^k variables are indicator variables for the number of treated links in person i's network. As noted above, K in our preferred model is 3+ (denoting that it includes "3 or more"), but the results hold for smaller and larger values for K as well. The second model is the following:

$$y_{iv} = \alpha_2 + \lambda_k \sum_{k=1}^{2} m_{iv}^k + \gamma_2 \mathbf{X}_i + \nu_{iv} \tag{5}$$

where m_i^k denotes the k^{th} moment of the distribution of signals in person i's network.

The baseline control variables include the share of treated households in a respondent's village who state that they know where to purchase the seeds, the number of seasons the respondent's household had used fertilizer in the four years before the time of the treatment, and the number of seasons they planted a hybrid variety (also in the past four years). These characteristics can be thought of as proxying for households' prior experience with improved technologies. Other household characteristics controlled for either proxy for household wealth or income (the size of the household's main maize field in acres, the PPI score, the type of rooftop, and a dummy for whether the household is in a microfinance group) or for individual characteristics (general risk attitudes, and a score indicating how well the participant understood an experimental game played before the network module). All regressions also include a linear control for total network size; the signal-regressions also control linearly for the number of treated links in a household's network.[32]

Most of the reported specifications are estimated with probits, except in the case of WTP, which is not a binary outcome variable and is therefore estimated using OLS.[33] The reported coefficients

[30] As Figure 5 showed, the distribution of farmers' rating of the performance of the sample packs is somewhat skewed to the left (i.e. positively skewed). Golec and Tamarkin (1998) discuss how bettors can be incorrectly identified as risk-loving if skewness in returns is not taken into account, but including the third moment of the distribution of signals in an individual's network does not change our results.

[31] In particular, we included a control for the number of network members who had had a negative experience, and we also controlled for whether or not an individual had *any* negative signals in her network.

[32] While the signal results are robust to controlling for the number of treated as dummy variables, the variance of the signals received by respondents who have zero or one treated links is not well-defined.

[33] Alternative results, available from the author upon request, estimate the WTP using a Tobit regression to account for possible censoring in the WTP. The rationale for estimating a Tobit is that we did not elicit a 'negative' WTP for those who ranked WSC hybrids below local varieties (which are assumed to cost 0 KES). It is therefore possible that those respondents would have wanted to actually *be paid* in order to be convinced to switch away from the local variety. In this case, their WTP for WSC hybrids would be negative. The estimates of network effects are slightly

are marginal effects when the results are obtained using probit regressions, and the regression coefficients from the OLS regressions.

In order to study the impact of farmers' own experience with the new technology, we have to assume that the performance of the new technology was unrelated to inherent farmer characteristics.[34] We can then include farmers' own perceived experimental gains in the basic regression.[35] While the assumption of exogeneity of the experience with the sample is strong, it is more plausible if we believe that farmers condition on their own ability and the fertility of their field when evaluating the technology. The question was phrased in order to encourage precisely such conditioning – farmers were asked what they believed they would have harvested had they planted what they normally plant under the same weather conditions and using the same inputs.

5.1.2 Heterogeneity and learning

To study the impact of heterogeneity on network effects, we exploit the fact that we have detailed soil quality data on treated farmers' fields. We single out the Cation Exchange Capacity (CEC) as a summary statistic of soil quality. It is widely recognized as the single best indicator of soil fertility (see for example Sanchez, 1976). Because it is predictive of fertility, farmers may know roughly where their neighbors fall in the distribution of CEC, but we don't expect farmers to be able to perfectly condition on it. In other words, we expect the assumptions of the model in Section 3 that farmers know the c_j's of their information neighbors to hold, but we don't expect them to know $f(c)$. While variation in factors on which an individual can adequately condition might actually help her learn,[36] whenever imperfectly observed individual characteristics are important determinants of neighbors' outcomes, social learning deteriorates (Munshi, 2004). Furthermore, the CEC is difficult to alter, and as such not affected by past management practices.

In our context, then, greater heterogeneity in a difficult-to-condition-on soil characteristic increases the likelihood that the signal in a farmer's network comes from an agent whose returns to the new hybrid is not very correlated with hers. We can also observe stark differences in the number of close neighbors (in terms of soil quality) that farmers identify: when asked to identify the farmers in the sample who have similar soils to theirs, more than half of our sample identified only one other farmer and 14% of the respondents stated they did not know whose soils are similar to theirs.

---

stronger with a Tobit than with OLS, but the Tobit does not converge for the indirectly treated households. Model convergence issues are common with MLE when the cell sizes for dummy variables are small, which here results from the few indirectly treated who answered the WTP module. The Tobit model for the treated shows the same patterns (larger coefficients, smaller constant) compared to OLS as the signal-model.

[34] Presumably, the main concern with including this variable would be omitted variable bias caused by better farmers both having positive experiences with the seeds and being more likely to have learned about the new technology/adopt a new technology.

[35] We also include the square of a farmer's own experience to allow for nonlinear effects. Excluding the square term does not affect the main results, suggesting either that nonlinearities are not very important or that this is not the correct form of nonlinearity.

[36] For an empirical example, Yamauchi (2007) finds that observed variation in schooling facilitates social learning about the returns to education

Another way of seeing this is by looking at the coefficient of variation (CV) in CEC. Since the levels of CEC vary substantially within the sample, the standard deviation would not provide a good measure of the heterogeneity in soil quality. We therefore normalize them using the CV. Figure 2 showed, there is significant variation in the CEC between and within villages, but also substantial variation between villages in how much intra-village variation there is. We use the CV of CEC in each village as our measure of heterogeneity. The CV in our sample villages ranges from just under 0.11 to just over 0.7, with an average of 0.32. To examine the relationship between heterogeneity and learning, we interact the information signals in person i's network, μ_i and σ_i^2, with the CV of soil quality in the village.[37]

5.2 Social network effects

Tables 4-7 show the results of our two information measures on each of the four outcome variables. Each regression table is divided into two panels: the top panel shows social-influence results (i.e. the number of treated links) while the bottom panel examines the mean and variance of the signals in an individual's network. All the estimates are obtained separately for the treated and the indirectly treated. Columns (1) and (4) show the most basic regression results, excluding covariates. Columns (2) and (5) add in covariates, and column (3) includes treated farmers' own perceived experimental gains.

Table 4 shows the network effects on farmers' familiarity with the new technology. While having more treated links has some effect for the treated farmers, the coefficients are halved when we control for household characteristics and become statistically indistinguishable from zero. This is not particularly surprising, since all the treated farmers received their own trial packets and should thus be familiar with the hybrid regardless of the number of treated links in their network. The number of treated links has some effect for the indirectly treated, but once we control for covariates the only significant network effect is the overall size of a person's network. There are several potential explanations for the impact of network size: more gregarious respondents may also be more curious or better informed; alternatively, "information passing" may occur through second-order links in the network, as observed by Banerjee et al. (2013). The authors find that non-participants – similar to the indirectly treated in our case – serve as important transmitters of information about microfinance, accounting for 1/3 of participation.

As for the information signals, Panel B of Table 4 shows that the variance of signals in a respondent's network increases their familiarity with the seeds. Since familiarity with the seed is unrelated to the perceived quality of the technology, this suggests that the buzz generated by a wide range of experiences with the trial packs affects individuals' likelihood of feeling informed about the seeds.

The network effects on WTP are positive and significant for both groups, although they disap-

[37] These results also hold if we use village fixed effects instead of directly controlling for each village's CV of CEC. Village fixed effects leave less variation and don't allow us to graph how the marginal impact of social networks change with village-level soil heterogeneity; we therefore choose the non-fixed effects model instead.

Table 4: Social network effects on farmer familiarity with WSC hybrids

Dep. variable: Familiar with WSC hybrid? (0/1)

Panel A -	**Treated**			**Indirectly treated**	
No. of treated links	1	2	3	4	5
1 treated in network	0.25*	0.16	0.65	0.27	0.073
	(0.1)	(0.2)	(0.7)	(0.2)	(0.2)
2 treated in network	0.23	0.12	0.73	0.36	0.12
	(0.2)	(0.2)	(0.7)	(0.2)	(0.2)
3+ treated in network	0.30**	0.17	0.78	0.39*	0.11
	(0.1)	(0.2)	(0.6)	(0.2)	(0.2)
Network size	0.0057	0.0091	0.036	0.0063	0.022***
	(0.007)	(0.006)	(0.03)	(0.01)	(0.009)
On-farm trial outcome			0.061		
			(0.2)		
Additional covars	NO	YES	YES	NO	YES
Observations	321	319	218	257	256

	Treated			**Indirectly treated**	
Panel B - Signal in nw	1	2	3	4	5
Avg. signal in nw.	0.15	0.14	0.12	0.24	0.21
	(0.1)	(0.1)	(0.10)	(0.2)	(0.2)
Variance of signal in nw.	0.28**	0.32**	0.36*	-0.00021	0.071
	(0.1)	(0.1)	(0.2)	(0.2)	(0.2)
No. in network	0.069	0.095	0.086	-0.024	0.024
	(0.05)	(0.06)	(0.06)	(0.08)	(0.07)
No. treated in network	-0.016	-0.078	0.0033	0.10	0.047
	(0.07)	(0.06)	(0.08)	(0.1)	(0.1)
On-farm trial outcome			0.060		
			(0.2)		
Additional covars	NO	YES	YES	NO	YES
Observations	292	291	202	224	224

Coeffs reported are probit marginal effects, with other explanatory vars held at sample mean

Standard errors clustered at the village level; * p<.1, ** p<.05, *** p<.01

See appendix for tables including all covariates. Regressions include interaction between no. of treated & network size.

Network definition used: individual j is in person i's corrected network (please see text for how correction was done) if person i listed them in *any* of the network questions.

Table 5: Social network effects on farmer WTP for WSC hybrids

Dep. variable: Willingness to pay for WSC hybrid (continuous)

Panel A -	**Treated**			**Indirectly treated**	
No. of treated links	1	2	3	4	5
1 treated in network	142.2*	56.5	75.3	392.9***	220.9***
	(80.9)	(93.1)	(128.0)	(42.7)	(68.3)
2 treated in network	165.3**	92.9	119.7	329.8***	165.6**
	(75.9)	(85.9)	(123.0)	(53.2)	(78.6)
3+ treated in network	119.2	54.6	41.9	323.9***	186.4***
	(70.8)	(75.6)	(108.3)	(31.5)	(58.7)
Network size	4.23	3.30	5.51	6.57	11.7
	(4.4)	(4.6)	(5.0)	(8.4)	(9.6)
On-farm trial outcome			49.0*		
			(24.6)		
Additional covars	NO	YES	YES	NO	YES
Observations	224	224	173	97	97
Adjusted R^2	0.003	0.057	0.079	0.006	0.119

	Treated			**Indirectly treated**	
Panel B - Signal in nw	1	2	3	4	5
Avg. signal in nw.	34.2	28.5	13.3	58.2*	61.9***
	(25.4)	(24.2)	(30.1)	(27.6)	(19.5)
Variance of signal in nw.	12.2	23.2	25.7	-46.4**	-46.0**
	(29.7)	(31.9)	(32.3)	(21.2)	(19.1)
No. in network	-1.96	-7.78	-10.2	1.32	6.19
	(8.9)	(8.5)	(10.9)	(14.3)	(17.3)
No. treated in network	-13.1	-7.55	-12.5	-17.7	-25.2
	(15.5)	(15.4)	(16.7)	(21.7)	(22.6)
On-farm trial outcome			0.060		
			(0.2)		
Additional covars	NO	YES	YES	NO	YES
Observations	292	291	202	224	224
Adjusted R^2	0.007	0.069	0.092	0.022	0.161

Standard errors clustered at the village level; * p<.1, ** p<.05, *** p<.01

See appendix for tables including all covariates. Regressions include interaction between no. of treated & network size.

Network definition used: individual j is in person i's corrected network (please see text for how correction was done) if person i listed them in *any* of the network questions.

Table 6: Social network effects on probability of planting a WSC hybrid

Dep. variable: Planted a WSC hybrid? (0/1)

Panel A -	**Treated**			**Indirectly treated**	
No. of treated links	1	2	3	4	5
1 treated in network	1.04***	0.97***	1.18***	0.44***	0.34***
	(0.2)	(0.2)	(0.2)	(0.1)	(0.07)
2 treated in network	0.93***	0.84***	1.04***	0.45***	0.32***
	(0.2)	(0.1)	(0.1)	(0.1)	(0.08)
3+ treated in network	0.95***	0.86***	1.07***	0.43***	0.31***
	(0.2)	(0.1)	(0.1)	(0.1)	(0.07)
Network size	0.0063	0.0098**	0.0098*	0.0034	0.0087*
	(0.005)	(0.004)	(0.006)	(0.005)	(0.005)
On-farm trial outcome			-0.0070		
			(0.03)		
Additional covars	NO	YES	YES	NO	YES
Observations	321	319	218	257	256

	Treated			**Indirectly treated**	
Panel B - Signal in nw	1	2	3	4	5
Avg. signal in nw.	0.38***	0.36***	0.48***	-0.14	-0.26
	(0.10)	(0.1)	(0.2)	(0.2)	(0.3)
Variance of signal in nw.	-0.069	-0.12	-0.22	0.32	0.44**
	(0.2)	(0.2)	(0.2)	(0.3)	(0.2)
No. in network	0.14**	0.14**	0.097	-0.036	0.027
	(0.06)	(0.06)	(0.08)	(0.09)	(0.1)
No. treated in network	0.024	0.019	0.10	-0.054	-0.058
	(0.1)	(0.1)	(0.2)	(0.1)	(0.2)
On-farm trial outcome			-0.16		
			(0.2)		
Additional covars	NO	YES	YES	NO	YES
Observations	292	291	202	224	224

Coeffs reported are probit marginal effects, with other explanatory vars gels at sample mean

Standard errors clustered at the village level; * p<.1, ** p<.05, *** p<.01

See appendix for tables including all covariates. Regressions include interaction between no. of treated & network size.

Network definition used: individual j is in person i's corrected network (please see text for how correction was done) if person i listed them in *any* of the network questions.

Table 7: Social network effects on probability of planting a non-WSC hybrid

Dep. variable: Planted a non-WSC hybrid? (0/1)

Panel A -	**Treated**			**Indirectly treated**	
No. of treated links	1	2	3	4	5
1 treated in network	-0.27	-0.40**	-0.37*	0.16	0.010
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
2 treated in network	-0.14	-0.26*	-0.30	0.21	-0.026
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
3+ treated in network	-0.094	-0.25*	-0.21	0.33*	0.060
	(0.2)	(0.1)	(0.2)	(0.2)	(0.1)
Network size	-0.00097	0.0037	0.0043	-0.034***	-0.0069
	(0.010)	(0.007)	(0.008)	(0.01)	(0.007)
On-farm trial outcome			0.074*		
			(0.04)		
Additional covars	NO	YES	YES	NO	YES
Observations	321	319	218	257	256

	Treated			**Indirectly treated**	
Panel B - Signal in nw	1	2	3	4	5
Avg. signal in nw.	0.13	0.16	0.12	0.17	0.18
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
Variance of signal in nw.	-0.074	-0.049	-0.014	-0.47***	-0.47***
	(0.1)	(0.2)	(0.2)	(0.1)	(0.2)
No. in network	-0.042	-0.0018	0.027	-0.13	-0.035
	(0.06)	(0.05)	(0.07)	(0.10)	(0.10)
No. treated in network	0.13*	0.11	0.11	0.24*	0.11
	(0.07)	(0.07)	(0.1)	(0.1)	(0.1)
On-farm trial outcome			0.24		
			(0.2)		
Additional covars	NO	YES	YES	NO	YES
Observations	292	291	202	224	224

Coeffs reported are probit marginal effects, with other explanatory vars gels at sample mean
Standard errors clustered at the village level; * p<.1, ** p<.05, *** p<.01
See appendix for tables including all covariates. Regressions include interaction between no. of treated & network size.
Network definition used: individual j is in person i's corrected network (please see text for how correction was done) if person i listed them in *any* of the network questions.

pear for the treated once we control for their own experience with the seeds and other covariates. As can be seen in columns (4) and (5) of Table 5, having a positive number of treated links increases an indirectly treated respondent's WTP by 200-400 Kenyan shillings (KES), depending on the specification and the number of treated links. For reference, 2-kg bag of the new hybrid costs roughly 350 KES.

We also find substantial impacts of the signal in the network: a one-standard-deviation increase of the average signal in an indirectly treated farmer's network (according to the estimates in column (5), Panel B of Table 5) results in a 15% increase in farmer WTP for the seeds. The precision of the signal also matters: a one-standard-deviation increase in the variance of a farmers' network information is associated with a 22% decrease in their WTP.

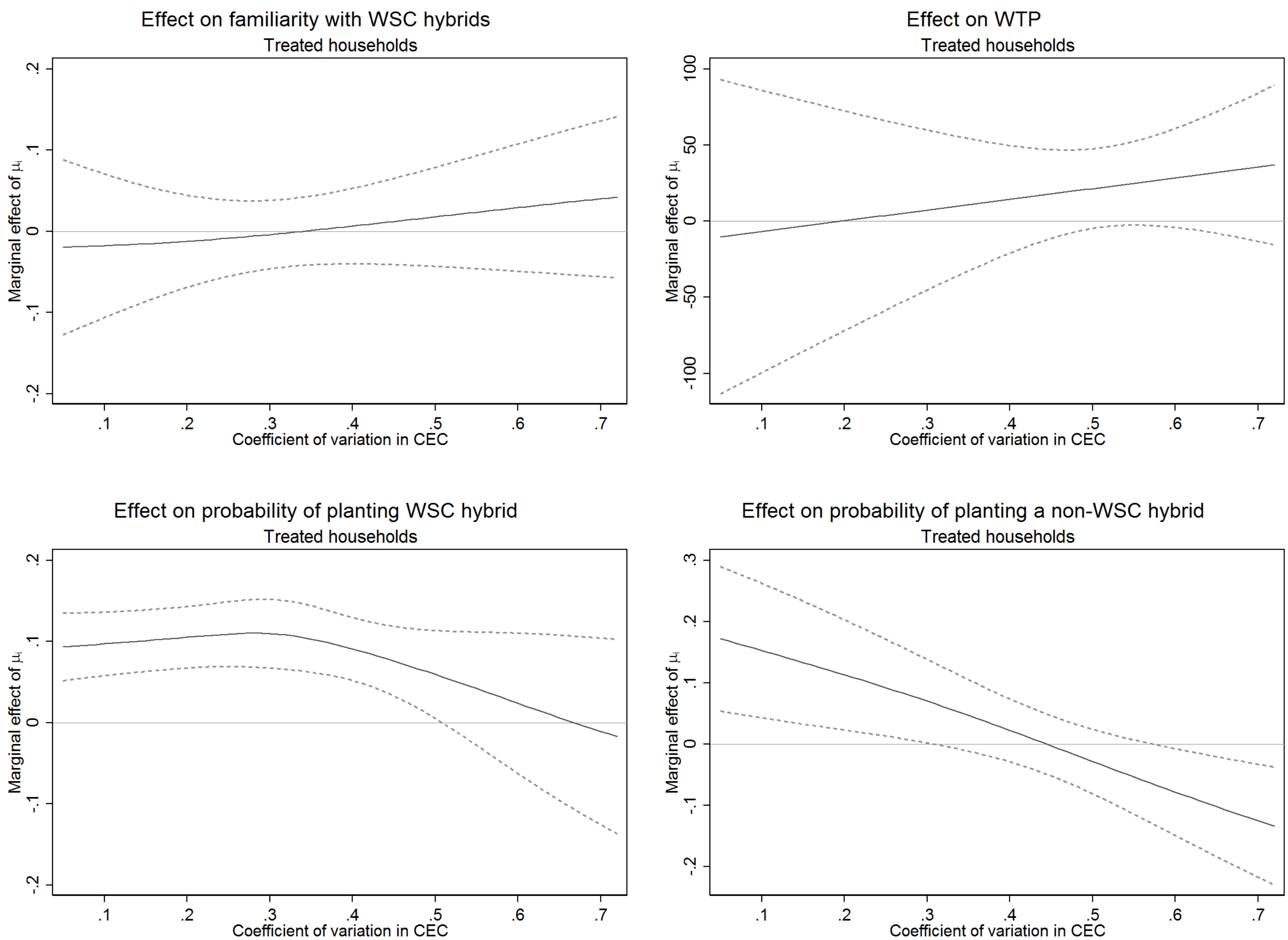
As for the probability of having planted a WSC hybrid in the current season, Table 6 shows that the impact of the number of treated links is substantial for both groups, while the average signal in the network seem to only matter for the treated – but the effects on the treated is large and positive. Surprisingly, a farmers' own on-farm trial experience does not have a significant impact on the probability of planting the seed, and this is not simply due to a high correlation between personal experience and the network information: a farmer's perceived experimental gains are not highly correlated with the mean and variance of their peers' experiences.

The social network effects on planting a different hybrid (Table 7), which we suggested earlier could be either null or negative, are in fact negative for the treated when it comes to the impact of the number of treated links and quite similar in absolute value to the coefficients in Table 6. This negative coefficient indicates that at least some of the adoption might be coming from a substitution away from other hybrids rather than a switch from local varieties. There appears also to be some spillover effect from the on-farm trial to other hybrids, since in column (2) we observe that a positive on-farm trial is associated with higher likelihood of planting a hybrid. For the indirectly treated, we see some effects in column (4) of the number of treated links, and a negative effect of the variance of the signal. In other words, having a less precise signal from one's network about the success of other farmers' on-farm experiences with WSC hybrids is associated with a lower likelihood of planting even another hybrid. This could potentially suggest that the indirectly treated believe they learned something about hybrids in general from signals in the network (rather than about the particular new hybrid that was introduced), and those with diffuse signals felt that they learned less.

5.3 Heterogeneity

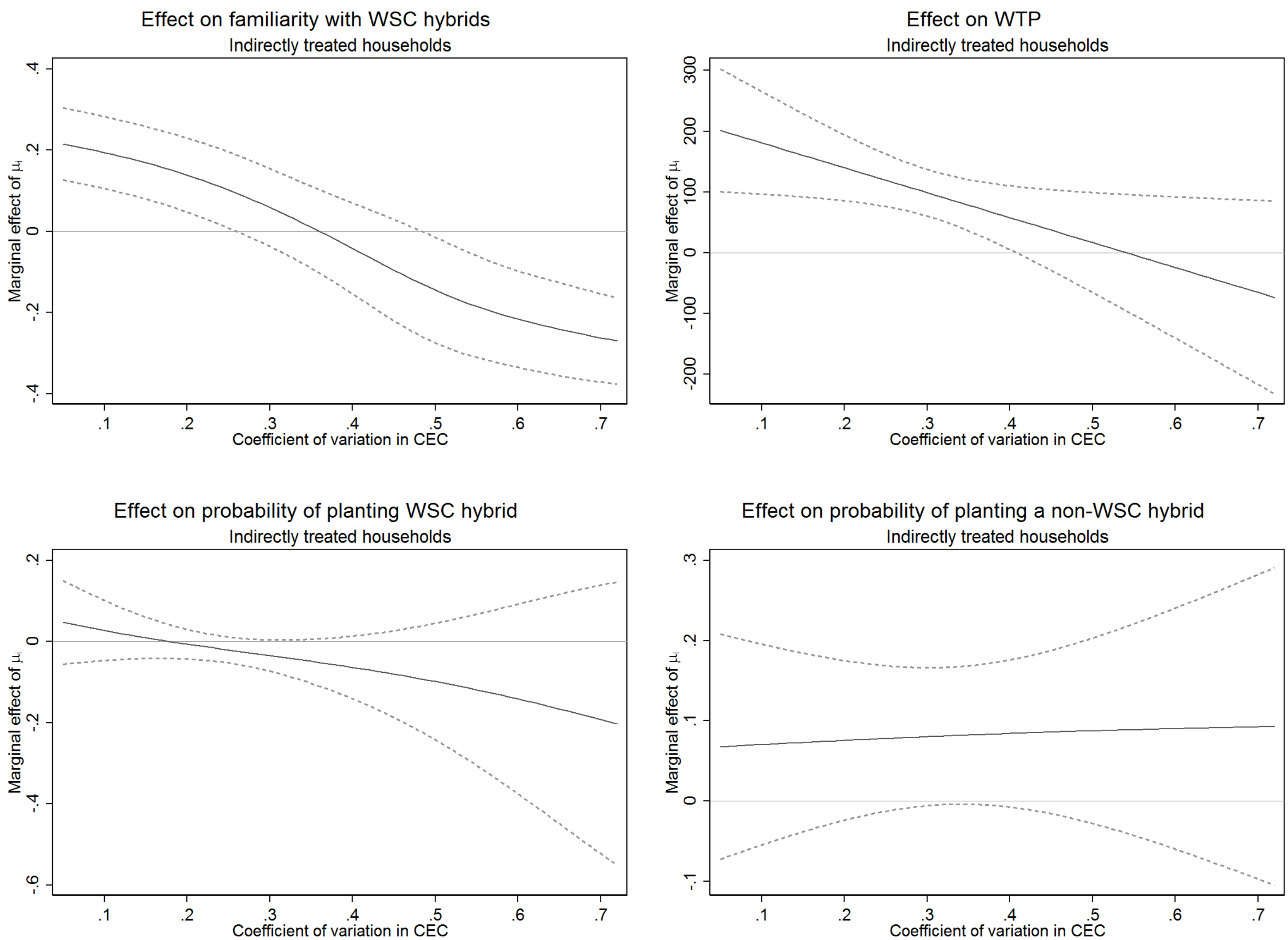
While the results above show that social network effects influence behavior in the sample villages, observing or talking to one's neighbor may be more or less useful depending on how similar these neighbors are along the dimensions that matter for the profitability of the technology. As Munshi (2004) shows, it is harder for individuals to learn from their network members about a technology that is sensitive to imperfectly observed characteristics when those characteristics vary in the population. Soil quality is an example of such a characteristic: crops respond to the fertility and

Figure 6: The marginal impact of the average signal in network on outcome variables against village soil quality heterogeneity (treated)



Note: Central solid lines show the marginal effects of the average signal in an individual's network; dashed lines show 90% confidence intervals.

Figure 7: The marginal impact of the average signal in network on outcome variables against village soil quality heterogeneity (indirectly treated)



Note: Central solid lines show the marginal effects of the average signal in an individual's network; dashed lines show 90% confidence intervals.

nutrient retention capabilities of the soils in which they grow, yet variation in soil quality within a village is not something that farmers can easily measure and control for when trying to learn from their fellow villagers' outcomes. Therefore, large variation in soil quality could negatively impact social learning.

We use precise measures of the fertility of the treated farmers' soils to compute the CV of soil quality in each village. We then interact the social network variables with the CV of Cation Exchange Capacity (CV_{CEC}), as described in more detail above in Section 5.1.2. Figures 6 and 7 show how the marginal impact of the average signal in individuals' networks on the four main outcome variables varies with the CV of CEC.[38]

While we don't see much of an effect of soil quality variability on the treated farmers' familiarity and WTP, the impact on actual adoption is quite stark. At low levels of soil quality variation, the average information signal in an individual's network positively influences an individual's likelihood of adopting WSC hybrids, but as the variation increases, the impact of the average signal decreases and becomes insignificant (and even slightly negative). The pattern for the indirectly treated is similar for all four outcome variables, and the point estimates are negative at high levels of variability for all the three outcomes that relate directly to WSC hybrids. This suggests that, absent a personal experience with the seed, farmers may actively disregard their peers experiences in highly heterogeneous environments.[39]

5.4 Robustness checks

This section outlines a few potential alternative mechanisms and robustness checks. First, we explore whether villages that are more heterogeneous are different along other dimensions than soil quality. Second, we discuss the implications of our networks being sampled (as opposed to capturing the complete network in a village). Third, we discuss some concerns about the potential endogeneity of the information signals in networks, which is not explicitly randomized.

5.4.1 Are heterogeneous villages different?

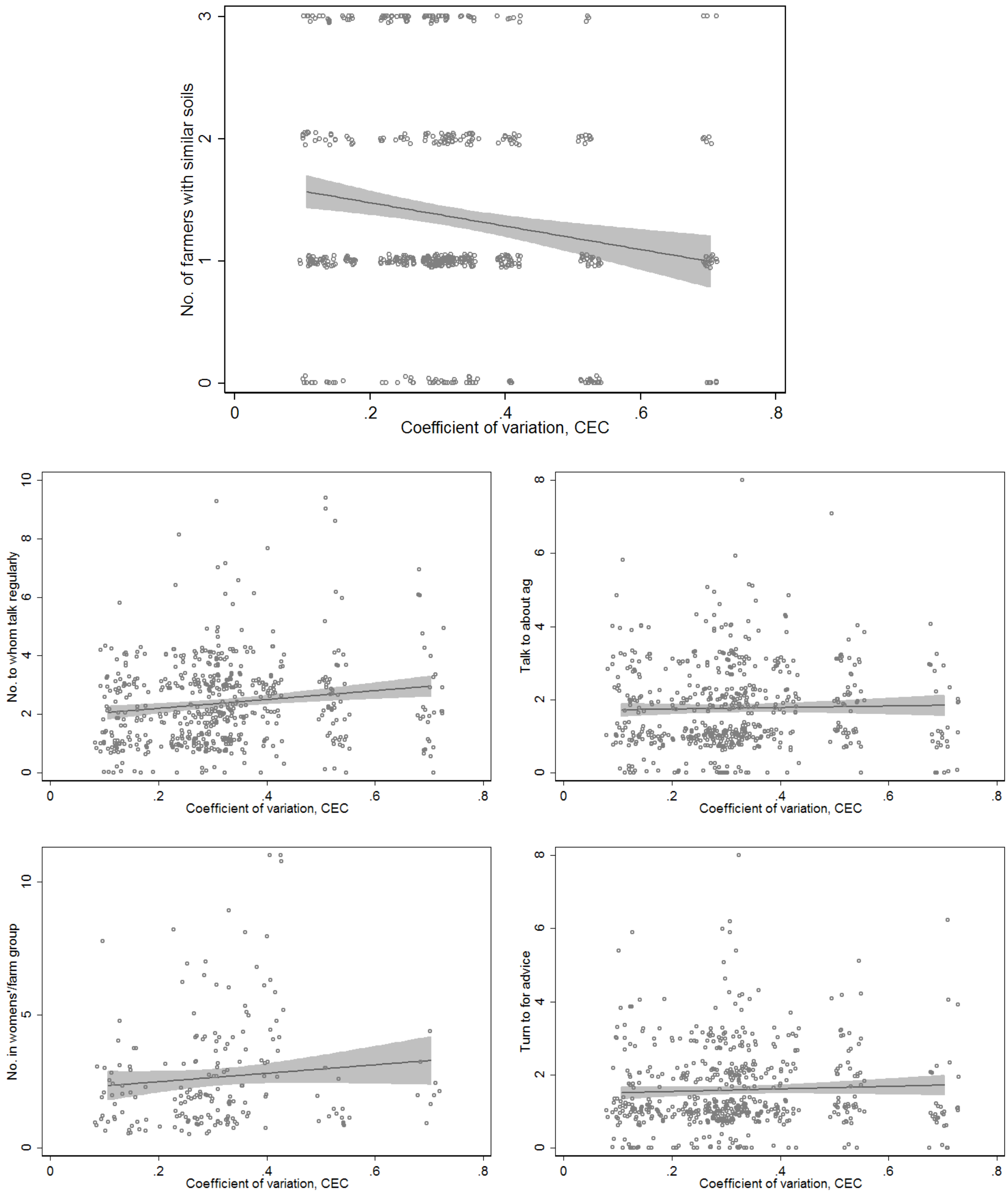
We may worry that social networks are weaker or more porous in heterogeneous villages, and that this drives the weaker learning results, While that would not be an uninteresting finding in and of itself, Figure 8 offers some suggestive evidence that it is not the case. The top graph shows a negative correlation between the number of farmers that a respondent listed as having similar soils to hers against the CV of CEC in her village.

This negative relationship does not appear for other network questions: for example, the figure also shows scatterplots of the number of contacts that *(i)* respondents talk to regularly, *(ii)* respondents talk to about agriculture, *(iii)* are in in respondents' farm groups, and *(iv)* respondents turn to for advice. In general, the relationships are either essentially flat or positive. Although not hard

[38]The range of CV_{CEC} corresponds to that observed in our villages.

[39]Tables 8-11 in Appendix D show the coefficient estimates of the moments of the distribution of signals at three different levels of CV of CEC for the four outcome variables.

Figure 8: Relationship between network characteristics and CV of soil quality



Note: Scatterplot jittered for clarity evidence, this indicates that farmers recognize the soil quality variability, and suggests that this is a plausible mechanism for the weaker social learning.

5.4.2 Sampled networks

The network measures used in this study are constructed from a partial sample of network nodes, which can give rise to econometric concerns. Chandrasekhar and Lewis (2011) discuss some of the implications of treating a sampled network as the true network of interest and propose analytic corrections for commonly used network statistics. The next version of our paper will correct for these biases, but for now we note that the illustrative examples that resemble our context suggest that our network effects are likely to be under-estimated due the partial sampling of the network. For example, Chandrasekhar and Lewis find that in a regression of a node's take-up decision on its neighbors' decisions, endogenous network effects have a 60% downward bias relative to the corrected estimate, and that they may even switch signs and become negative. Moreover, regression coefficients in several of their specifications are not significantly different from zero at conventional levels when using the raw sampled data but are significantly different when applying the reconstruction estimator.

5.4.3 Endogeneity of signals

While the number of treated links in a person's network is exogenous in this setting due to the randomization, the experiences of an individual's peers may be correlated with her own and as such subject to the standard identification problems in the network literature. This can arise due to, for example, correlated unobservables or common shocks. A common approach the literature is to use the characteristics of an individual's friends as instruments for her own characteristics (Bramoulle et al., 2009). In this context, since we ware concerned about the distribution of signals in your network being endogenous, this is the characteristic that we would want to instrument for. In order to do so, need a strong relationship between friends' signals and treated people's signals. In fact, the correlation between person i's network signals and her peers' signals do not provide a strong enough first stage to use them as instruments. F-statistics on the first stage regression hover around 7. While precluding the use of an IV strategy, it should simultaneously alleviate the concerns about endogeneity in the first place.

6 Conclusion

This paper uses data from a randomized experiment to examine what farmers learn from their social networks about a new agricultural technology in western Kenya and how this affects their behavior. Unlike another study in the region (Duflo et al., 2011), which finds little evidence of informational spillovers about agricultural technologies, we find strong evidence of information passing as well as peer effects in adoption.

The researchers on the contrasting study (Duflo et al., 2011) suggest that the observed lack of network effects is due to farmers not talking much to each other about agriculture, but that it is unlikely due to heterogeneity in soil quality across farms. We find that farmers mention talking to each other quite frequently and that heterogeneity in soil quality matters significantly. We explicitly measure soil quality and find that higher variation in fertility within a village decreases social network effects by making it harder for farmers to learn from each other.

A better understanding of the complexity that farmers face when learning about new technologies is key to understanding why some innovations diffuse more slowly than would be socially optimal. Our results indicate that farmers react to the presence of heterogeneity by relying less on information from their peers when making agricultural decisions. This finding highlights the importance of recognizing the complexity that farmers face when designing extension policies.

While farmers can learn from each other, the more variable the environment, the more important learning-by-doing becomes. On the one hand, policy makers can take these findings into account by ensuring that input recommendations and extension services bear local conditions in mind. On the other hand, while we have shown that social network effects can help diffuse information about agricultural technologies, promoting individual learning may be optimal in particularly heterogeneous regions. In the case of hybrids, this could be achieved by subsidizing learning or by making samples of seeds available to farmers for on-farm trials.

References

Babcock, P. S. and J. L. Hartman (2010): "Networks and workouts: treatment size and status specific peer effects in a randomized field experiment," Working Paper 16581, National Bureau of Economic Research.

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

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The Diffusion of Microfinance," *Science*, 341, 1236498.

Besley, T. and A. Case (1994): "Diffusion as a learning process: Evidence from HYV cotton," Working paper.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2011): "Chapter 18 - Identification of Social Interactions," in *Handbook of Social Economics*, ed. by A. B. a. M. O. J. Jess Benhabib, North-Holland, vol. Volume 1, 853–964.

Bramoulle, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

Breza, E. (Forthcoming): "Field experiments, social networks, and development," in *Oxford Handbook on the Economics of Networks*, ed. by Y. Bramoulle, A. Galeotti, and B. Rogers.

Cai, J., A. De Janvry, and E. Sadoulet (2015): "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 7, 81–108.

Carter, M. R., R. Laajaj, and D. Yang (2014a): "Subsidies and the Persistence of Technology Adoption: Field Experimental Evidence from Mozambique," Working Paper 20465, National Bureau of Economic Research.

Carter, M. R., T. J. Lybbert, M. K. Mathenge, and E. Tjernstrom (2014b): "Spurring technological innovation and poverty reduction? Evaluating the impact of a new seed market actor in Kenya," BASIS Brief 2014-04, BASIS Assets and Market Access Innovation Lab.

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

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): "Individual risk attitudes: Measurement, determinants, and behavioral consequences," *Journal of the European Economic Association*, 9, 522–550.

Duflo, E., M. Kremer, and J. Robinson (2011): "Learning by Doing but not Learning from Others: Experimental Evidence from Maize Farmers in Kenya," Working paper, Washington, D.C.

Foster, A. D. and M. R. Rosenzweig (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103, 1176–1209.

Golec, J. and M. Tamarkin (1998): "Bettors Love Skewness, Not Risk, at the Horse Track," *Journal of Political Economy*, 106, 205–225.

Jones, C. and J. Jacobsen (2005): "Plant nutrition and soil fertility," Report, Montana State University Extension Service.

Kenya Plant Health Inspectorate Service (2014): "National crop variety list - Kenya," Report, Kenya Plant Health Inspectorate Service.

Ligon, E. and E. Sadoulet (2008): "Estimating the effects of aggregate agricultural growth on the distribution of expenditures," Background Paper, World Bank, Washington, D.C.

Maertens, A. (2014): "Who cares what others think (or do)? Social learning, social pressures and imitation in cotton farming in India," Working Paper.

Magnan, N., D. J. Spielman, T. J. Lybbert, and K. Gulati (2013): "Leveling with Friends: Social Networks and Indian Farmers Demand for a Resource Conserving Technology," Working Paper 01302, IFPRI.

Manski, C. F. (1993): "Dynamic choice in social settings: Learning from the experiences of others," *Journal of Econometrics*, 58, 121–136.

Mathenge, M. K., M. Smale, and J. Olwande (2012): "The Impact of Maize Hybrids on Income, Poverty, and Inequality among Smallholder Farmers in Kenya," in *Innovations in impact assessment of agricultural research: Theory and practice*, Meetings of the International Association of Agricultural Economists (IAAE), Foz do Iguacu, Brazil.

Miguel, E. and M. Kremer (2004): "Worms: identifying impacts on education and health in the presence of treatment externalities," *Econometrica*, 72, 159–217.

Munshi, K. (2004): "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 73, 185–213.

Oster, E. and R. Thornton (2012): "Determinants of Technology Adoption: Peer Effects in Menstrual Cup Take-up," *Journal of the European Economic Association*, 10, 1263–1293.

Sanchez, P. (1976): *Properties and Management of Soils in the Tropics*, New York: John Wiley & Sons, Ltd.

Smale, M. and J. Olwande (2014): "Demand for maize hybrids and hybrid change on smallholder farms in Kenya," *Agricultural Economics*, 45, 409–420.

Suri, T. (2011): "Selection and Comparative Advantage in Technology Adoption," *Econometrica*, 79, 159–209.

Tittonell, P., B. Vanlauwe, M. Corbeels, and K. E. Giller (2008): "Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya," *Plant and Soil*, 313, 19–37.

World Bank (2008): "World development report 2008: Agriculture for development," Working Paper.

Yamauchi, F. (2007): "Social learning, neighborhood effects, and investment in human capital: Evidence from Green-Revolution India," *Journal of development Economics*, 83, 37–62.

Appendix A - Network questions

1. Which of these households are your relatives? (share the same grandfather)
2. Which of these households do you talk to on a regular basis? (Unspecified topic / about anything)
3. Which of these households are in your microfinance group? (Conditional on household being in a microfinance group)
4. Which of these households are in your women's group / farming group? (Conditional on household being in a women's group/farming group)
5. Which of these households' fields do you WALK/BIKE BY regularly?
6. Which of these households' fields do you live closest to?
7. Which of these households do you discuss agriculture with?
8. Out of the people you discuss agriculture with (automated based on previous selection), who did you start communicating with about maize farming in the past 12 months?
9. Which households do you talk to about agriculture *more than once a year*? (Filtered based on who they discuss agriculture with)
10. Which households do you talk to about agriculture *more than once a month*? (Filtered based on who they talk to more than once a year)
11. Which households do you talk to about agriculture *more than once a week*? (Filtered based on who they talk to more than once a month)
12. Which households would you turn to if you needed advice about growing maize?
13. Out of all the households, do you know what maize seed they planted these long rains? (Select if the answer is yes)
    (a) For the selected, indicate what seed they planted
14. Out of all these households, do you know what maize seed is their referred seed? (Select if the answer is yes)
    (a) For the selected, indicate what seed they prefer
15. Which of these households is the most similar to you as a farmer? (In terms of farming practices, etc.)
16. Which of these households have the most similar soils to yours?

17. With which of these households have you spoken to about the WSC seed samples that were distributed last year?
18. Which of these households would recommend WSC hybrids?

Appendix B - Proof of proposition 1

As stated, soil quality is drawn from the positive real half-line, i.e. $c_i,\, c_1, ..., c_n > 0$. We wish to examine the conditional distribution of $f(c_i)$ given $f(c_1), ..., f(c_n)$. Let $\left(c_{(1)}, ..., c_{(n)}\right)$ represent the order statistics of the sample of soil qualities $(c_1, ..., c_n)$.

We then need to examine three different cases:

1. Farmer i's soil quality is the lowest in the village, i.e. $0 < c_i < c_{(1)}$
2. Farmer i has neighbors whose soil quality lie above and below hers, i.e. $c_{(j)} < c_i < c_{(j+1)}$
3. $c_i > c_{(n)}$

In order to show that the variance of farmer i's posterior belief is an increasing function of σ_c^2, we need the means and variances of the conditional distribution of $f(c_i)$ in each of these cases. Cases 1 and 2 are Brownian bridges, and the result in Case 3 comes from the definition of Brownian motion. I include the derivations below for completeness.

<u>Case 1</u>, $0 < c_i < c_{(1)}$, we can add and subtract the term $\frac{c_i}{c_{(1)}} f\left(c_{(1)}\right)$ to obtain the following equality:

$$f(c_i) = \frac{c_i}{c_{(1)}} f\left(c_{(1)}\right) + \left(f(c_i) - \frac{c_i}{c_{(1)}} f\left(c_{(1)}\right)\right)$$

where the term $f(c_i) - \frac{c_i}{c_{(1)}} f\left(c_{(1)}\right)$ is independent of the sequence of observations $(f(c_1), ..., f(c_n))$. In fact, for each $k = 1, 2, ..., n$ we have

$$\begin{aligned} E\left(f(c_i) - \frac{c_i}{c_{(1)}} f\left(c_{(1)}\right)\right) f(c_{(k)}) &= Ef(c_i)f(c_{(k)}) - \frac{c_i}{c_{(1)}} Ef(c_{(k)})f(c_{(k)}) \\ &= c_i - \frac{c_i}{c_{(1)}} c_{(1)} = 0 \end{aligned}$$

where the last equality uses the fact that the covariance of the Brownian motion: $E\left[f(t)f(s)\right] = \min(s, t)$.

So the conditional distribution of $f(c_i)$ is normal with mean $\frac{c_i}{c_{(1)}} f(c_{(1)})$ and variance equal to

$$\begin{aligned} E\left(f(c_i) - \frac{c_i}{c_{(1)}} f\left(c_{(1)}\right)\right)^2 &= Ef(c_i)^2 - 2\frac{c_i}{c_{(1)}} Ef(c_i)f(c_{(1)}) + \frac{c_i^2}{c_{(1)}^2} Ef(c_{(1)})^2 \\ &= c_i - 2\frac{c_i^2}{c_{(1)}} + \frac{c_i^2}{c_{(1)}} \\ &= \frac{c_i(c_{(1)} - c_i)}{c_{(1)}} \end{aligned}$$

For case 2, $c_{(j)} < c_i < c_{(j+1)}$, we can use the representation

$$f(c_i) = \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} f(c_{(j)}) + \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} f\left(c_{(j+1)}\right)$$
$$+ \left(f(c_i) - \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} f(c_{(j)}) - \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} f\left(c_{(j+1)}\right)\right)$$

where the term $f(c_i) - \frac{c_{(j+1)}-c_i}{c_{(j+1)}-c_{(j)}} f(c_{(j)}) - \frac{c_i-c_{(j)}}{c_{(j+1)}-c_{(j)}} f\left(c_{(j+1)}\right)$ is independent of the sequence of observations $(f(c_1), ..., f(c_n))$. In fact, for each $k \le j$, we have

$$E\left(f(c_i) - \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} f(c_{(j)}) - \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} f\left(c_{(j+1)}\right)\right) f(c_{(k)}) =$$
$$= Ef(c_i)f(c_{(k)}) - \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} Ef(c_{(j)})f(c_{(k)}) - \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} Ef(c_{(j+1)})f(c_{(k)}) =$$
$$= c_i - \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} c_{(j)} - \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} c_{(j+1)} = 0$$

Therefore, the conditional distribution of $f(c_i)$ is normal with mean

$$\frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} f(c_{(j)}) + \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} f(c_{(j+1)})$$

and variance equal to

$$E\left(f(c_i) - \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} f(c_{(j)}) - \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} f\left(c_{(j+1)}\right)\right)^2 =$$
$$= E\left(f(c_i) - \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} f(c_{(j)}) - \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} f\left(c_{(j+1)}\right)\right) f(c_i) =$$
$$= c_i - \frac{c_{(j+1)} - c_i}{c_{(j+1)} - c_{(j)}} c_{(j)} - \frac{c_i - c_{(j)}}{c_{(j+1)} - c_{(j)}} c_{(i)} =$$
$$= \frac{\left(c_{(j+1)} - c_i\right)\left(c_i - c_{(j)}\right)}{c_{(j+1)} - c_{(j)}}$$

with the first equality stemming from the orthogonality of $f(c_{(j)})$ and $f(c_{(j+1)})$.

For case 3, $c_i > c_{(n)}$, we can use the representation

$$f(c_i) = f(c_{(n)}) + (f(c_i) - f(c_{(n)}))$$

The term $f(c_i) - f(c_{(n)})$ is independent of the sequence of observations $(f(c_1), ..., f(c_n))$ since increments of Brownian motion are independent. Therefore, the conditional distribution of $f(c_i)$ is normal with mean $f(c_{(n)})$ and variance equal to

$$E(f(c_i) - f(c_{(n)}))^2 = c_i - c_{(n)}$$

In sum, the posterior variance of $f(c_i)$ given $(f(c_1), ..., f(c_n))$ equals

$$V_{posterior} = \sum_{j=1}^{n} \frac{\left(c_{(j+1)} - c_i\right)\left(c_i - c_{(j)}\right)}{c_{(j+1)} - c_{(j)}} 1_{c_{(j-1)}<c_i<c_{(j)}} + \left(c_i - c_{(n)}\right) 1_{c_i > c_{(n)}}, \tag{6}$$

where for convenience we normalize $c_{(0)} = 0$.

The next step is to take expectations of Eq. (6) and to show that it is an increasing function of σ_c^2. To do this, we note that we can assume the following identities

$$c_i = \sigma_c \xi_i,\ c_1 = \sigma_c \xi_1, ...,\ c_n = \sigma_c \xi_n$$

for ξ, $\xi_1, ..., \xi_n$ independent standard normal variables.

Further consider the ordered sample $\xi_{(1)}, ..., \xi_{(n)}$. By definition of ordered statistics, $\xi_{(1)}$ is the smallest of all the $\xi_j = \frac{c_j}{\sigma_c}$ in the sample. Therefore, $\sigma_c \xi_{(1)} = c_j$ is the smallest of all the c_j's in the sample, i.e. $\sigma_c \xi_{(1)} = c_{(1)}$. The same holds for the other elements of the ordered sample. So the ordered samples are also related by $c_{(i)} = \sigma_c \xi_{(i)}$. So then we can write the following expression for the posterior belief:

$$E(V_{posterior}|c_i,\ c_1, ..., c_n > 0) =$$

$$= E\left(\sum_{j=1}^{n} \frac{\left(c_{(j+1)} - c_i\right)\left(c_i - c_{(j)}\right)}{c_{(j+1)} - c_{(j)}} 1_{c_{(j-1)}<c_i<c_{(j)}} + \left(c_i - c_{(n)}\right) 1_{c_i > c_{(n)}}|c_i,\ c_1, ..., c_n > 0\right) =$$

$$= E\Bigg(\sum_{j=1}^{n} \frac{\sigma_c^2 \left(\xi_{(j+1)} - \xi_i\right)\left(\xi_i - \xi_{(j)}\right)}{\sigma_c \left(\xi_{(j+1)} - \xi_{(j)}\right)} 1_{\sigma_c\xi_{(j-1)}<\sigma_c\xi_i<\sigma_c\xi_{(j)}} +$$

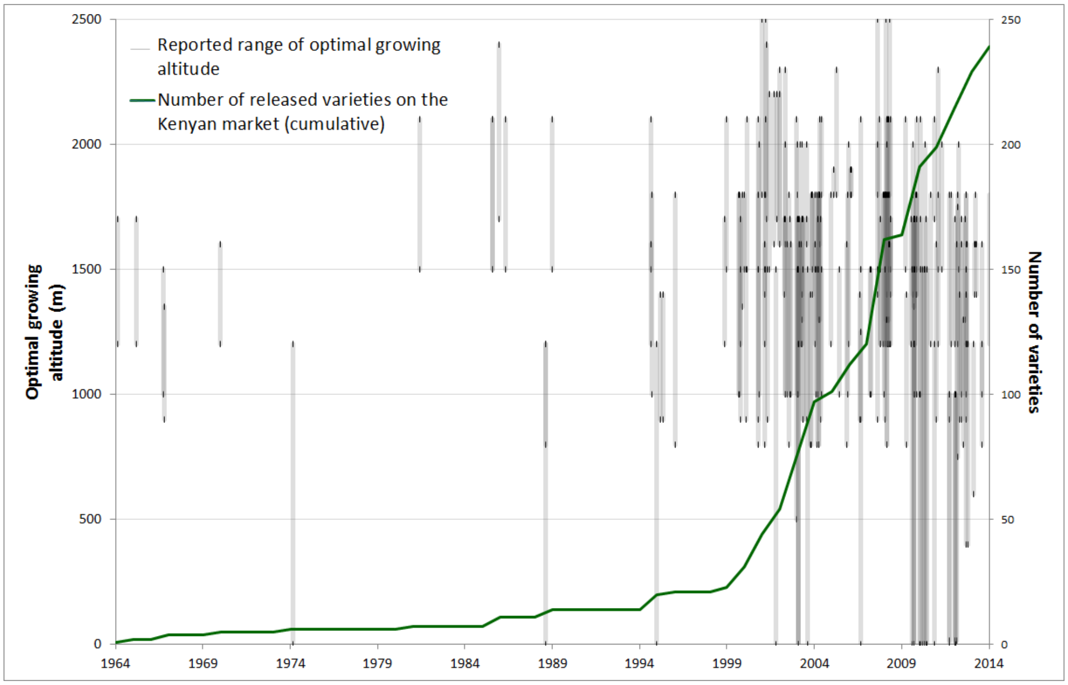
$$\sigma_c \left(\xi_i - \xi_{(n)}\right) 1_{\sigma_c\xi_i > \sigma_c\xi_{(n)}}|\xi_i,\ \xi_1, ..., \xi_n > 0)\Bigg)$$

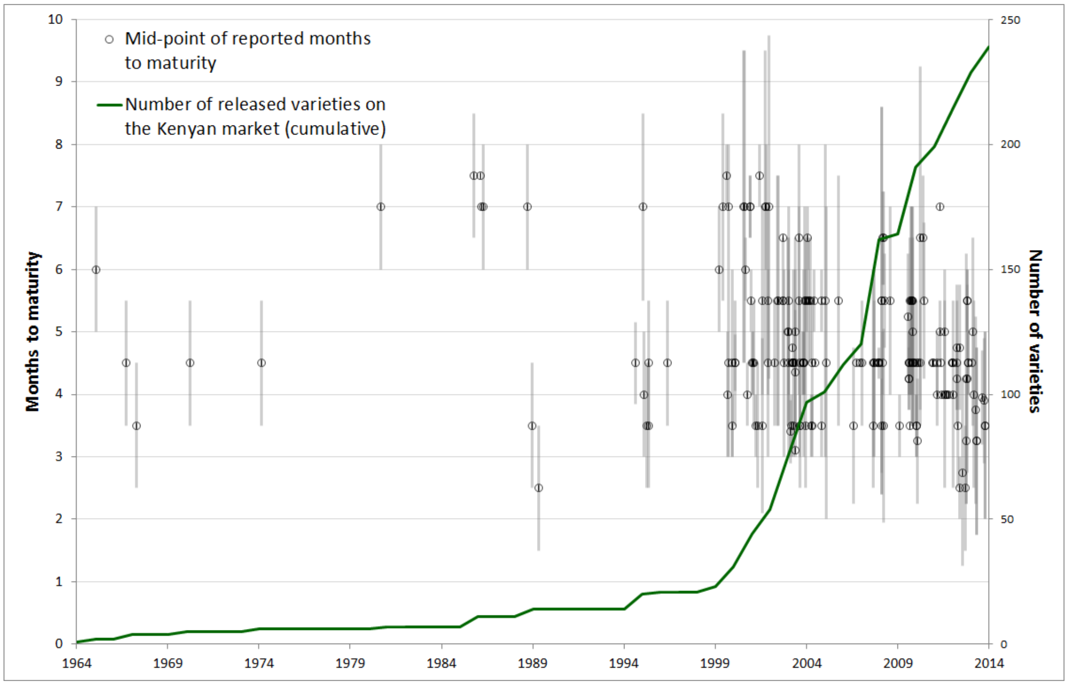
$$= \sigma_c \Bigg(\sum_{j=1}^{n} E\left(\frac{\left(\xi_{(j+1)} - \xi_i\right)\left(\xi_i - \xi_{(j)}\right)}{\left(\xi_{(j+1)} - \xi_{(j)}\right)} 1_{\xi_{(j-1)}<\xi_i<\xi_{(j)}}|\xi_i,\ \xi_1, ..., \xi_n > 0\right) +$$

$$E\left(\left(\xi_i - \xi_{(n)}\right) 1_{\xi_i > \xi_{(n)}}|\xi_i,\ \xi_1, ..., \xi_n > 0\right)\Bigg)$$

So the expected value of the farmer's posterior belief is a linear function of σ_c, with a positive slope, since all the terms in the expectations are positive.

Appendix C - Maize variety information





Appendix D

Table 8: Impact of heterogeneity on social network effects: familiarity with WSC hybrids

(Dependent variable: Indicator for respondent is familiar with WSC hybrid)

	Treated			**Indirectly treated**		
Percentile of CV_{CEC}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}
Mean(signal)	-0.0044	-0.00020	0.035	0.19***	0.057	-0.26***
	(0.05)	(0.03)	(0.06)	(0.05)	(0.06)	(0.07)
Var(signal)	0.0023	0.062**	0.18	-0.13	-0.029	0.22
	(0.06)	(0.03)	(0.1)	(0.1)	(0.04)	(0.2)
Network size	0.024*	0.024*	0.021**	0.0036	0.0045	0.0038
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
No. of treated in nw	-0.034	-0.033*	-0.030*	0.025	0.031	0.027
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Observations	291	291	291	224	224	224

In both panels: standard errors in parentheses; s.e.'s clustered at the village level

* p<.1, ** p<.05, *** p<.01

Table 9: Impact of heterogeneity on social network effects: : WTP

(Dependent variable: Willingness to pay for WSC hybrid)

Estimated using OLS	**Treated**			**Indirectly treated**		
Percentile of CV_{CEC}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}
Mean(signal)	1.23	11.3	33.1	174.1***	97.8***	-66.6
	(51.5)	(29.7)	(29.5)	(45.4)	(22.1)	(87.8)
Var(signal)	74.3*	15.4	-111.5*	-165.7*	-56.2***	179.9
	(42.3)	(22.0)	(62.5)	(79.9)	(19.3)	(118.1)
Network size	0.54	0.54	0.54	14.6	14.6	14.6
	(8.4)	(8.4)	(8.4)	(12.5)	(12.5)	(12.5)
No. of treated in nw	3.40	3.40	3.40	1.09	1.09	1.09
	(13.5)	(13.5)	(13.5)	(18.4)	(18.4)	(18.4)
Observations	212	212	212	92	92	92

In both panels: standard errors in parentheses; s.e.'s clustered at the village level

* p<.1, ** p<.05, *** p<.01

Table 10: Impact of heterogeneity on social network effects: planting a WSC hybrid

(Dependent variable: Indicator for planted WSC hybrid in main season, 2014)

	Treated			**Indirectly treated**		
Percentile of CV_{CEC}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}
Mean(signal)	0.10***	0.11***	-0.014	0.021	-0.035	-0.19
	(0.03)	(0.03)	(0.07)	(0.04)	(0.02)	(0.2)
Var(signal)	-0.061*	-0.022	0.17**	0.026	0.032	0.040
	(0.04)	(0.02)	(0.08)	(0.03)	(0.1)	(0.1)
Network size	0.015*	0.022**	0.030*	0.0072	0.0063	0.0081
	(0.008)	(0.01)	(0.02)	(0.007)	(0.005)	(0.006)
No. of treated in nw	-0.013	-0.019	-0.026	0.0055	0.0048	0.0062
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Observations	291	291	291	224	224	224

In both panels: standard errors in parentheses; s.e.'s clustered at the village level

* p<.1, ** p<.05, *** p<.01

Table 11: Impact of heterogeneity on social network effects: planting a non-WSC hybrid

(Dependent variable: Indicator for planted a non-WSC hybrid in main season, 2014)

	Treated			**Indirectly treated**		
Percentile of CV_{CEC}	5^{th}	50^{th}	95^{th}	5^{th}	50^{th}	95^{th}
Mean(signal)	0.15**	0.072*	-0.13**	0.072	0.081	0.093
	(0.06)	(0.04)	(0.06)	(0.07)	(0.05)	(0.1)
Var(signal)	0.028	-0.026	-0.13	-0.12	-0.099***	-0.052
	(0.07)	(0.03)	(0.1)	(0.08)	(0.04)	(0.1)
Network size	-0.0086	-0.0097	-0.0085	-0.011	-0.011	-0.011
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
No. of treated in nw	0.024	0.027	0.024	0.024	0.024	0.024
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Observations	291	291	291	224	224	224

In both panels: standard errors in parentheses; s.e.'s clustered at the village level

* p<.1, ** p<.05, *** p<.01