ECONOMIC GROWTH CENTER

YALE UNIVERSITY

P.O. Box 208269
New Haven, Connecticut 06520-8269
http://www.econ.yale.edu/~egcenter/

CENTER DISCUSSION PAPER NO. 817

LEARNING ABOUT A NEW TECHNOLOGY: PINEAPPLE IN GHANA

Timothy G. Conley
Northwestern University

and

Christopher R. Udry
Yale University

July 2000

Note: Center Discussion Papers are preliminary materials circulated to stimulate discussions and critical comments.

This research has received financial support from the NSF (grants SBR-9617694 and SES-9905720), Clair Brown Fellowship, International Food Policy Research Institute, Institute for the Study of World Politics, World Bank Research Committee, Fulbright Program, and Social Science Research Council.

The authors have benefitted from the advice of Richard Akresh, Ana Fernandez, Garth Frazer, Markus Goldstein, Ethan Ligon, Charles Manski, Francesca Molinari, Stephen Morris, Chris Taber, Giorgio Topa, and seminar participants at a number of seminars. The data used in this paper were collected by Ernest Appiah, Robert Ernest Afedoe, Patarick Selorm Amihere, Esther Aku Sarquah, Kwabena Moses Agyapong, Esther Nana Yaa Adofo, Michael Kwame Arhin, Margaret Harriet Yeboah, Issac Yaw Omane, Peter Ansong-Manu, Ishmaelina Borde-Koufie, Owusu Frank Abora, and Rita Allotey while the second author was a visiting scholar at the Institute of Statistical, Social and Economic Research at the University of Ghana. The authors are of course responsible for any errors in this paper.

LEARNING ABOUT A NEW TECHNOLOGY:
PINEAPPLE IN GHANA

Timothy G. Conley
Northwestern University

and

Christopher R. Udry
Yale University

Abstract

This paper investigates the role of social learning in the diffusion of a new agricultural technology in a developing country: Ghana. We use unique data on farmers' communication patterns to define each individual's information neighborhood, the set of others from whom he might learn. Our empirical strategy is to test whether farmers change their input decisions to align with those of their neighbors who were successful in previous periods. We present evidence that farmers adopt successful neighbors' practices, conditional on many potentially confounding factors including the physical proximity of plots, credit arrangements, clan membership, and soil characteristics.

KEY WORDS: Social Learning, Technology, Innovation

JEL CLASSIFICATION: 031, 012, 013

1 Introduction

The transformation of technology is fundamental to the development process. For a new technology to be adopted by an agent, particularly in agriculture, it must be adapted to the circumstances faced by that agent. Its characteristics usually will not be transparent to the new user (Evenson and Westphal (1995)). Consequently, an investment in learning about the new technology is associated with its adoption. If there are multiple adopters of the new technology in similar circumstances, as is often the case with an innovation in agriculture, then the process of learning about the new technology may be *social*. New users of the technology may learn its characteristics from each other.

The role of social learning in promoting growth and technology diffusion has been featured in the endogenous growth literature (Romer (1986); Lucas (1988); Aghion and Howitt (1998)). It is also an integral part of current practice in agricultural research and extension systems in developing countries. New technologies are introduced either by farmers' own experimentation or through formal sector intervention and the process of social learning encourages their diffusion (Bindlish and Evenson (1997); Rogers (1995)). Only recently, however, have economists made efforts to measure the quantitative importance of learning from others.[1]

In this paper we investigate learning about a new agricultural technology by farmers in the Akwapim South district of Ghana. Over the last decade, an established system of maize and cassava intercropping for sale to urban consumers has begun a transformation into intensive production of pineapple for export to European markets (Obeng (1994)). This transformation of the region's farming system involves the adoption of a set of new technologies, including intensive use of agricultural chemicals not used in the previous system. These farmers' need to learn new techniques, combined with our access to micro-level data during this period (from 1996 to 1998), make this a good setting in which to study social learning.

Measuring the extent of social learning is difficult for two major reasons. First, the set of neighbors from whom an individual can learn is difficult to define. Second, even with a proper definition of this set, distinguishing learn-

[1] In contrast, there is a long tradition of empirical studies by economists of the adoption of new technologies in agriculture. Griliches (1957) is the seminal work. For reviews see Feder et al (1985) and Evenson and Westphal (1995). This important literature, however, does not isolate the role of learning processes from other determinants of adoption.

ing from other phenomena that may give rise to similar observed outcomes is problematic. In the absence of learning, individuals may still act like their neighbors as a result of interdependent preferences, technologies, or because they are subject to related unobservable shocks.

In data commonly used by economists, direct information on information interconnections is typically unavailable.[2] Consequently, economic investigations of the process of social learning have typically made assumptions that relate observed relationships between individuals - such as geographical proximity - to unobserved flows of information. This set of assumptions is critical for the measurement of the extent of social learning, but can rarely be tested because of data limitations.[3] Foster and Rosenzweig (1995), for example, provide tabulations indicating that 'friends and neighbors' are an important source of information about fertilizer use, but are limited to using village averages as the relevant information set for social learning.

We have access to unusually rich data that allows us to address the concerns of neighbor definition more directly. Our approach draws on the classic work by Coleman *et al* which related adoption of new antibiotics to the network of social interconnections between the doctors. We collected detailed information on who individuals know and talk to about farming. Hence we follow Coleman *et al* by defining information links between agents using responses to questions about which other agents they turn to for information (Coleman *et al* (1957, p. 254)). Rogers (1995) and Birkhaeuser *et al* (1991) provide valuable surveys of research that describes and characterizes the set of neighbors from whom agents learn about new innovations in a wide variety of settings.

Once neighborhoods are defined, the identification of learning is still a formidable problem. The classic problem of omitted variables prevents us from inferring that learning effects must be present simply from observations on, say, the diffusion process of a new technology. The fact that a farmer

[2]Exceptions include Woittiez and Kapteyn (1998) and Kapteyn (2000) who use individuals' responses to questions about their 'social environments' to describe their reference groups.

[3]In many investigations of learning in developing country agriculture, the reference group is taken to be all farmers in the village (Foster and Rosenzweig (1995), Besley and Case (1994), Yamauchi K (2000)). Munshi and Myaux (1998) take exceptional care in the construction of reference groups for social learning by using external evidence on communication barriers arising from religion. See Manski (1993) for a concise discussion of the importance of reference group designations in identification of endogenous social effects.

is more likely to adopt a new technology soon after his neighbors have done so might be a consequence of some unobserved variable that is spatially and serially correlated, rather than learning. We have unusually good data for reducing the extent of this problem. Our data contains detailed geographic and soil information as well as information on credit and family relationships, allowing us to control for many potentially confounding factors.

However, as Foster and Rosenzweig (1995) point out, even without omitted variable problems, mimicking behavior (in which farmers simply copy their neighbors' behavior) may produce the same alignment of actions as learning. They advocate investigating the impact of neighbors' actions upon a farmer's productivity to help distinguish learning from mimicry. They investigate the influence of neighbors' adoption decisions on an individual's profits. We also rely on observations of farmers' profitability to distinguish learning from mimicry. We use profit measures to define the set of neighbors that a farmer uses to guide his behavior.

Our identification problem can be thought of as a special case of the general problem of identification in social interactions models studied by Manski (1993, 1997), Moffitt (1999), Brock and Durlauf (1999) and others.[4] This literature is concerned with the problem of inferring whether an individual's behavior is influenced by the behavior of those in his neighborhood or reference group. Manski (1993) demonstrates that identification depends crucially on the relationship between the variables that define reference groups and those that directly influence outcomes. With cross sectional data, prospects for identification are best when these two variables are 'moderately' related: the conditional expectations of direct variables are not linear in the group variables, they are not functionally dependent, nor are they independent. With panel data, prospects for identification improve, subject to the important caveat that the assumptions regarding the timing of social interactions must be properly specified. Our strategy for identifying learning effects relies on intertemporal comparisons of farmers' actions and the outcomes of previous actions of their neighbors. We investigate whether farmers change their input decisions to align with those of their neighbors with similar circumstances who were previously successful, earning high profits. The timing of responses to neighbors' actions arises naturally from the staggered revelation of information from preceding plantings.

[4]See Brock and Durlauf (1999) for a survey of the literature on social interactions models.

We model farmers' learning using a target input model.[5] Each harvest opportunity gives the farmer an observation on output for a given amount of inputs, and thus reveals information about the optimal input distribution. Learning by doing can be modeled in this framework as updating beliefs using information from one's own harvests; social learning is updating given information about neighbors' actions and harvests. Our primary method to test for social learning is to estimate how farmers' input decisions respond to the actions and outcomes of their neighbors. From plot-level farming data, we know the inputs used and output harvested by each farmer, and thus the outcome of each 'experiment' with the new technology by each respondent. We use knowledge/communication data to trace the impact of the information revealed by each experiment on the future input decisions of other farmers who are in the 'information neighborhood' of the cultivator who conducted the experiment. Our data allows us to control for many confounding factors including common shocks due to physical proximity of plots, credit arrangements, clan membership, and soil characteristics.

We examine social learning about the use of fertilizer in pineapple cultivation. We find strong effects of 'successful experiments' with fertilizer in the information neighborhood of a farmer on that farmer's future innovations in fertilizer use. Specifically, we find that a farmer (say, farmer A) increases (decreases) his use of fertilizer after other farmers in his information neighborhood achieve higher than median profits when using more (less) fertilizer than he did.[6] This relationship holds for a variety of different definitions of 'information neighborhood.' Moreover, we can distinguish this apparent learning effect from many of the confounding effects that often obscure the interpretation of correlations between the actions of agents. Conditional on successful experiments in the information neighborhood farmer A, there is no relationship between farmer A's innovations in fertilizer use and the cultivation practices of farmers who cultivate physically nearby plots, or with

[5] Target input models have been widely used in both theoretical and empirical work. A partial list includes Prescott (1972), Wilson (1975), Jovanovic and Nyarko (1994), Foster and Rosenzweig (1995). The target input model is not always appropriate, because as commonly specified it does not permit disadoption when the technology is revealed to be unprofitable. Besley and Case (1994) propose an alternative approach to avoid this restriction. For pineapple in the study area, there has been no disadoption.

[6] We use the male pronoun to refer to farmers because of the extreme gender division of labor associated with pineapple in this area: of 93 pineapple farmers in the sample, only 8 are women.

whom farmer A has financial ties. In principle, our empirical method could distinguish learning from mimicking behavior by contrasting responses to actions of successful and unsuccessful neighbors. Unfortunately, we do not have sufficient variation in our data to rule out mimicry as an alternative explanation for this finding. We do find, however, that farmers substantially increase their profits by modifying their fertilizer use in the direction of the amount of fertilizer used in recent successful experiments in their information neighborhood.

The remainder of this paper is organized as follows. In Section 2 we present the simple learning model that motivates our empirical specifications. Section 3 describes the empirical setting and our data. Section 4 provides a description of our empirical model and Section 5 describes our results. Section 6 offers a brief conclusion.

2 Learning Model

This section describes a simple model of learning about optimal inputs that we use to guide our empirical work. The basic form of this model is that farmers know the production technology up to the distribution of an optimal or target input. Suppose that each farmer has the following technology and that each operates it on one plot, for simplicity in notation. On farmer i's plot, a single input (fertilizer) $f_{i,t}$ is chosen at time t and the following period output $q_{i,t+1}$ is produced. The amount of output is random because the *ex post* optimal level of the input is random, given by a 'target input' variable $\theta_{i,t+1}$ that for simplicity we take to be IID across agents and time.[7] We assume the loss associated with suboptimal inputs is quadratic with the following form:

$$q_{i,t+1} = \lambda - \left(f_{it} - \theta_{i,t+1}\right)^2 .$$

The distribution of the target input shock is assumed to be multinomial with three distinct points of support $x < y < z$ with probabilities p_x, p_y, and $(1 - p_x - p_y)$, respectively. Farmers are assumed to know the form of this production function and the distribution of the target input shock and have only the probabilities p_x and p_y to learn.

[7] In our empirical work, we do allow for dependence across observations.

We model farmers' knowledge about $[p_x\ p_y]$ in a Bayesian framework so that they have a subjective prior distribution over values of these probabilities. The conjugate prior distribution for the multinomial is the Dirichlet distribution, characterized in this example by three parameters: $T_{i,x}, T_{i,y}$, and $T_{i,z}$ which we take to be integers for simplicity.[8] So the prior distribution we will work with can be written as:

$$h_i(p_x, p_y) \propto {p_x}^{T_{i,x}-1} {p_y}^{T_{i,y}-1} (1 - p_x - p_y)^{T_{i,z}-1}. \tag{1}$$

The farmer's problem is to choose input $f_{i.t}$ given his beliefs. In this target input model, there are no strategic nor experimental motivations for a farmer to vary his use of inputs. Farmers are assumed to know the form of the production function, therefore the same information is revealed by any level of input use, so the optimal strategy is to choose $f_{i,t}$ to maximize profits.[9]

For simplicity of exposition, we assume that inputs are costless. The assumption of quadratic loss from having suboptimal inputs implies that the optimal choice of inputs is the expectation of θ with respect to his subjective prior distribution. The expected value of p_x under the distribution (1) is $\left[\frac{T_{i,x}}{T_{i,x}+T_{i,y}+T_{i,z}}\right]$. The expected value of p_y is $\left[\frac{T_{i,y}}{T_{i,x}+T_{i,y}+T_{i,z}}\right]$, and so on. Therefore, farmer i's subjective expectation of θ given these priors is:

$$E_{io}(\theta_{i,1}) = x\left[\frac{T_{i,x}}{T_{i,x}+T_{i,y}+T_{i,z}}\right] + y\left[\frac{T_{i,y}}{T_{i,x}+T_{i,y}+T_{i,z}}\right] + z\left[\frac{T_{i,z}}{T_{i,x}+T_{i,y}+T_{i,z}}\right].$$

This quantity will be the optimal input choice for farmer i with beliefs given by (1).

We focus our attention on three simple examples where farmers' learning about $[p_x\ p_y]$ corresponds to updating of their beliefs using Bayes' rule. The first example is a two-farmer village where both farmers are each other's

[8] The reader might think of these parameters as the number of occurances of x, y, and z, respectively, from a set of previous draws from the multinomial distribution of θ that have been observed by farmer i. However, in our behavioral model of updating, we assume that the farmers take these priors as primative, uninformative and common knowledge. Farmer B, for example, is assumed to know farmer A's prior, and does not interpret it as evidence of previous draws from the distribution of θ. Morris (1995) provides a useful discussion of models with heterogeneous priors.

[9] This is conditional on cultivating the new crop. There are strategic considerations in the adoption decision, see Foster and Rosenzweig (1995), Besley and Case (1994).

information neighbor. We also suppose that each farmer observes without error all inputs and outcomes for himself as well as his information neighbor. The second example illustrates the impact that limited observability would have on updating rules in the same two-farmer, one-neighborhood village. Finally, the last example illustrates the process of social learning when networks matter, that is, when not all farmers are information neighbors of each other. With many neighborhood structures, observations of neighbors' actions contain information beyond that generated by observations of the outcomes of their trials themselves.

Example 1

Suppose there are two farmers A and B, each has one plot, and each is the other's neighbor. Assume that each farmer has full information *ex post* about the other's harvest and input choices. At time period t, farmer A can observe $q_{B,t}$ and the associated input $f_{B,t-1}$ in addition to his own input and output, $f_{A,t-1}$ and $q_{A,t}$. With his knowledge of the production technology, farmer A can deduce the realization of the optimal input shocks for both plots: $\theta_{A,t}, \theta_{B,t}$. Suppose they start out in period zero with the different priors defined by (1). Farmer $i's$ period zero actions will be $f_{i,0} = E_{i,0}(\theta_{i,1})$. Say, for example, the *ex post* target input shocks in period one are both equal to $x : \theta_{A,1} = \theta_{B,1} = x$. In this case, both farmers update their beliefs to reflect the knowledge that two draws of x have been observed and their posterior means for p_x will be $\frac{T_{i,x}+2}{(T_{i,x}+2)+T_{i,y}+T_{i,z}}$ and for $p_y : \frac{T_{i,y}}{(T_{i,x}+2)+T_{i,y}+T_{i,z}}$. Thus both farmers reduce their use of f in period 1:

$$\begin{aligned} f_{i,1} &= E_{i,1}(\theta_{i,2}) = x\left[\frac{T_{i,x}+2}{(T_{i,x}+2)+T_{i,y}+T_{i,z}}\right] + \\ & y\left[\frac{T_{i,y}}{(T_{i,x}+2)+T_{i,y}+T_{i,z}}\right] + z\left[\frac{T_{i,z}}{(T_{i,x}+2)+T_{i,y}+T_{i,z}}\right] < f_{i,0} \\ for\; i &\in \{A, B\}. \end{aligned}$$

Suppose for example farmer B had initial priors that led him to believe that the expected value of θ was low relative to the initial beliefs of farmer A, so $f_{B,0} < f_{A,0}$. Given that the target input shocks in period one were both equal to x, farmer B is relatively successful ($q_{B,1} > q_{A,1}$) and we observe in period one both farmers reducing their input use.

The variance of each farmer's subjective distributions is inversely proportional to the number of draws as their beliefs become concentrated on the true p_x and p_y. Farmer i's initial subjective variance of p_x with beliefs represented by (1) is

$$\frac{T_{i,x}(T_{i,y}+T_{i,z})}{(T_{i,x}+T_{i,y}+T_{i,z})^2\,(T_{i,x}+T_{i,y}+T_{i,z}+1)}$$

and those for p_y and p_z are analogous. Upon gaining the information that $\theta_{A,1}=\theta_{B,1}=x$, the variances of farmer i's beliefs about p_x are:

$$\frac{(T_{i,x}+2)\,(T_{i,y}+T_{i,z})}{((T_{i,x}+2)+T_{i,y}+T_{i,z})^2\,((T_{i,x}+2)+T_{i,y}+T_{i,z}+1)}$$

These updated subjective variances reflect the increases in information and will vanish as the number of observed experiments approaches infinity.

This example illustrates that with full information, neighbors' experiments are just as informative as a farmer's own experiments and so they have a symmetric impact upon beliefs. It illustrates the principle that with full information, a farmer's beliefs in any period depend only upon his priors and the number of realizations of x, y and z from previous trials by anyone in his information neighborhood. Neither the identity of the farmer nor the ordering of the realizations has any effect on posterior beliefs. Finally, the example makes the simple and robust point that a farmer will tend to adjust his input use in the direction of the level of inputs used by relatively successful farmers in his information neighborhood.

Example 2.

In this example we illustrate the consequence of limited observability of output in a two farmer village where each is the other's neighbor. Suppose that there is full information about one's own inputs and output as well as one's neighbor's inputs. However, farmer A observes only an indicator of whether the harvest of farmer B's plot was good, above a level $q^* : 1\left(q_{B,t} > q^*\right)$. Values of q^* such that $(\lambda - q^*)$ is interior to the range of $\left(f_{B,t} - \theta_{B,t+1}\right)^2$ will reveal limited information about the realized target input shock. For example, suppose q^* and $f_{B,t}$ are such that

$$\left(f_{B,t}-z\right)^2 < (\lambda - q^*) < \left(f_{B,t}-y\right)^2 \;\; < \;\; \left(f_{B,t}-x\right)^2 .$$

In this case, observing that $q_{B,t} > q^*$ implies that the target shock was z but $q_{B,t} < q^*$ implies only that the target shock was not z. Thus, the information revealed to farmer A is that the target input for farmer B is either z or in the set $[x, y]$. Suppose that farmer A starts out with the same priors as above and receives only the information: $\theta_{B,1} \in [x, y]$. The likelihood of this event is:

$$L(data|p_x, p_y) = (p_x + p_y)$$

Therefore, farmer A's posterior is:

$$h_i(p_x, p_y|data) \propto (p_x + p_y)\, p_x{}^{T_{A,x}-1} p_y{}^{T_{A,y}-1} (1 - p_x - p_y)^{T_{A,z}-1}. \tag{2}$$

Straightforward but tedious calculation reveals the posterior means for A of p_x,p_y, and p_z :

$$\begin{aligned}
E_A(p_x) &= \left[\begin{array}{c} \frac{T_{A,x}}{T_{A,x}+T_{A,y}} \left(\frac{1+T_{A,x}}{1+T_{A,x}+T_{A,y}+T_{A,z}}\right) + \\ \frac{T_{A,y}}{T_{A,x}+T_{A,y}} \left(\frac{T_{A,x}}{1+T_{A,x}+T_{A,y}+T_{A,z}}\right) \end{array}\right] \\
E_A(p_y) &= \left[\begin{array}{c} \frac{T_{A,x}}{T_{A,x}+T_{A,y}} \left(\frac{T_{A,y}}{1+T_{A,x}+T_{A,y}+T_{A,z}}\right) + \\ \frac{T_{A,y}}{T_{A,x}+T_{A,y}} \left(\frac{1+T_{A,y}}{1+T_{A,x}+T_{A,y}+T_{A,z}}\right) \end{array}\right] \\
E_A(p_z) &= \left(\frac{T_{A,z}}{1+T_{A,x}+T_{A,y}+T_{A,z}}\right).
\end{aligned} \tag{3}$$

The impact of limited observability can be seen by comparing the posterior means in (3) to those that would result from farmer A observing that $\theta_{B,1} = x$:

$$\begin{aligned}
E_A(p_x) &= \frac{1+T_{A,x}}{1+T_{A,x}+T_{A,y}+T_{A,z}} \\
E_A(p_y) &= \frac{T_{A,y}}{1+T_{A,x}+T_{A,y}+T_{A,z}} \\
E_A(p_z) &= \frac{T_{A,z}}{1+T_{A,x}+T_{A,y}+T_{A,z}}.
\end{aligned} \tag{4}$$

Thus observing the event that $\theta_{B,1}$ is x or y results in a posterior mean for p_A that is a convex combination of the posterior mean that would have resulted if $\theta_{B,1} = x$ and that if $\theta_{B,1} = y$, and the weights correspond to the relative weights of p_x and p_y under the prior: $T_{A,x}$ and $T_{A,y}$, respectively. The posterior mean for p_z is the same as it would have been given exact knowledge of whether $\theta_{B,1}$ equaled x or y. The updated $E_{A,1}(\theta_{A,2})$ is a linear combination of x, y, and z with weights given by the posterior means in expression (3) and in general will be distinct from that generated from observing the exact value of $\theta_{B,1}$.

Suppose again that farmer B starts with a prior belief that the expected value of θ is low relative to the initial beliefs of farmer A. Hence, $f_{B,0} < f_{A,0}$. In this example as well, given that the target input shocks in period one were both equal to x, farmer B is relatively successful ($q_{B,1} > q_{A,1}$). After observing that $\theta_{B,1} \in \{x, y\}$, the imperfectly-informed farmer A now has a lower expected value of θ. In period one we observe farmer A modifying his fertilizer use in the direction of the lower $f_{B,0}$ chosen by the relatively successful farmer B in the initial period, although not as far as in the full information case of Example 1.

Example 3.

It is possible that learning occurs through social networks rather than in the context of the collective experiment illustrated in Examples 1 and 2. The following example considers the implications for learning when information is shared within overlapping neighborhoods. Suppose there are three farmers A,B and C with the following information neighborhood structure: A has only B for a neighbor, C has only B for a neighbor, and B has both A and B for neighbors. Assume as in the first example that each farmer has full information about his neighbors' trials.

Consider the optimal actions of farmers A and B if they update with only the information in the inferred θ signals of their neighbors. The three farmers may have different priors and thus different choices of $f_{i,0}$ (to be concrete, suppose $f_{A,0} > f_{B,0} > f_{C,0}$). Suppose the realizations of θ in the period 1 trials are $\theta_{A,1} = \theta_{B,1} = \theta_{C,1} = x$. After period 1, farmer A updates his beliefs as follows:

$$\begin{aligned} E_{A,1}(\theta_{A,2}) &= x\left[\frac{T_{A,x}+2}{(T_{A,x}+2)+T_{A,y}+T_{A,z}}\right]+ \\ &\quad y\left[\frac{T_{A,y}}{(T_{A,x}+2)+T_{A,y}+T_{A,z}}\right]+z\left[\frac{T_{A,z}}{(T_{A,x}+2)+T_{A,y}+T_{A,z}}\right]. \end{aligned}$$

Farmer B observes the outcomes of all three trials, so he updates to:

$$\begin{aligned} E_{B,1}(\theta_{B,2}) &= x\left[\frac{T_{B,x}+3}{(T_{B,x}+3)+T_{B,y}+T_{B,z}}\right]+ \\ &\quad y\left[\frac{T_{B,y}}{(T_{B,x}+3)+T_{B,y}+T_{B,z}}\right]+z\left[\frac{T_{B,z}}{(T_{B,x}+3)+T_{B,y}+T_{B,z}}\right]. \end{aligned}$$

As in the previous examples, the optimal input choice for each farmer is to set $f_{i,t} = E_{i,t}(\theta_{i,t+1})$. Note that farmer A knows all the determinants of $f_{B,1} = E_{B,1}(\theta_{B,2})$ with the exception of $\theta_{C,1}$. He will be able to determine the value of $\theta_{C,1}$ when he observes farmer B's action in period one.[10] Therefore, if Farmer A were to use only the information contained in $\{\theta_{A,1}, \theta_{B,1}, \theta_{A,2}, \theta_{B,2}\}$, he would not have an optimal update of the probabilities because he would omit the information generated by $\theta_{C,1}$. In general, when learning occurs through networks, observation of neighbors' optimal decisions contains a source of information that is not captured by their target input shock histories. Moreover, the particular history of realizations of θ now matters because farmer A must model farmer B's updating process in order to deduce farmer B's observation of $\theta_{C,1}$. If the network structure were to be extended, for example by adding farmer D who has only farmer C as a neighbor, farmer A would have to extend the depth of his memory by an additional period in order to deduce $\theta_{D,1}$ and $\theta_{C,2}$.

It can be seen that the problem faced by (for example) farmer A of deducing the realizations of $\theta_{i,t}$ for farmers i not in the information neighborhood of farmer A becomes very complex for extended neighborhood structures. We make no attempt to model this process explicitly in the empirical work that follows. Instead, we examine the simpler implication of each of these models that farmers' beliefs and observable actions are updated in response to their observation of new information from neighbors' experiments. In each of these examples, if a farmer's beliefs are different from those of his information neighbor (and thus he chooses a different level of input application) and if that neighbor is relatively successful, then the farmer tends to update his beliefs and actions in the direction of those of his neighbor.[11] Successful experiments, in this sense, attract imitation.

[10] Given any $\{T_{i,x}, T_{i,y}, T_{iz}\}$, this statement holds for generic choice of $\{(x, y, z) \in \mathbb{R}^3_+ | x < y < z\}$.

[11] In Example 3, it will occasionally be the case that a farmer updates his fertilizer use away from that of a successful neighbor. In this example, suppose that z is much larger than y, and that farmer C realized $\theta_{C,1} = z$, rather than x as assumed in the text. Then after A deduces from $B's$ period 1 action that $\theta_{C,1} = z$, it could be the case that A updates his period 2 action to increase, rather than decrease his fertilizer use. However, even in this example on average farmers will update their fertilizer use to reflect that of their succesful neighbors as fertilizer use by each converges towards the optimal level of $p_x x + p_y y + (1 - p_x - p_y)z$.

3 Empirical Setting

This section describes the data we use in our empirical work. First, we discuss the measures of farmers' knowledge and communication with each other that we use to define neighborhoods. Then we describe the economic and agronomic data that we use to describe farmers' learning. The section ends with a discussion of the salient features of our neighborhood data.

The data are drawn from a two-year survey of approximately 240 households in southern Ghana. A fairly comprehensive set of individual and agronomic data was collected. As some of the data was quite sensitive, we limited the size of the sample in order to maintain close oversight of the interview process. The sample was constructed in two stages. The process began with the purposive selection of four 'villages' near the towns of Nsawam and Aburi.[12] This region is the center of the recent growth of intensive vegetable cultivation in the Eastern Region. The second stage was a random sample of married individuals: 60 couples (or triples, when there are two wives) were chosen by a simple random sample in each village. Two enumerators lived in or near each village and interviewed each respondent in 15 rounds at intervals of approximately six weeks.

In addition to the data on pineapple production, communication, knowledge and social networks described in 3.1 and 3.2 we make use of data on the characteristics of farmers.[13] Wealth is defined as the value of the non-land assets held by the farmer at the start of the survey period. The clan indicator variables denote membership in a particular *abusua*, or matrilineal clan. The religion indicator denotes membership in a particular charismatic church. A table of summary statistics is included in the appendix.

3.1 Communication and Knowledge Data

One of our main innovations is that we are able to use survey data to directly define 'information neighborhoods.' Most of the literature in economics on learning about agriculture has been forced to rely on strong assumptions regarding the flow of information in communities. In contrast, we can base our

[12] We use data from the three of these 'villages' where pineapple is farmed. We will refer to the units as 'villages' although, in fact, only two are legally villages. The other two locations are a pair of adjacent villages and a village with a set of outlying hamlets.

[13] A detailed description of survey procedures, copies of the survey instruments and the data archive are available at http://www.econ.yale.edu/~cru2/ghanadata.html.

measures of information availability on direct data about conversations between individuals or other indicators that individuals share knowledge about farming. Specifically, we have two classes of information links measures: one set is based on data on interactions between selected pairs of individuals; the second on a roster of contacts for each respondent.

To concisely describe the definitions of our metrics let $M = \{1, ..., m\}$ be the set of pineapple farmers in a village and i and j be typical members of this set. Let $l_{ij} \in \{0, 1\}$ describe the relationship between any two farmers in the village. We say that i and j are linked if $l_{ij} = 1$, and the neighborhood of i is defined as $N_i = \{j | l_{ij} = 1\}$.

Our 'pairs' information measures are based on samples of plots and of individuals. Each sample respondent was questioned about a random sample (without replacement) of seven other individuals in the same village, and with three other predetermined individuals who appear to be focal in the village. Links for each respondent are defined according to his responses to specific questions about the relationship with the selected persons. In addition, each respondent was matched with a random sample of six recently-harvested pineapple *plots* in the same village. Links are defined between the respondent and the cultivator of the relevant plot.

The samples of individuals produced responses to the question: "Have you ever gone to ___ for advice about your farm?". In this case, we define $l_{ij} = 1$ if either i responded 'yes' to this question about j or if j responded 'yes' to this question about i. In 11% of the matches, one of the farmers had gone to the other for advice. We use responses to this question as our benchmark definition of information neighbors because during the field research it appeared reliably-answered and it is transparently related to the learning process under study.

The sample of pineapple plots yields responses to the following questions. "Do you know ___?". In this case, we define $l_{ij} = 1$ if either i responded 'yes' to this question about j or if j responded 'yes' to this question about i. In more than 80% of the matches respondents knew the cultivator of the selected plot. If the respondent knows the cultivator, we ask "Do you know that s/he has a pineapple farm at [insert description of plot]?". In this case, we define $l_{ij} = 1$ if either i responded 'yes' to this question about any of j's pineapple plots or if j responded yes to this question about any of i's pineapple plots. The pineapple plot was known in just over 30% of total matches (including those where 'do you know __?' was answered no). If the respondent knows about the pineapple plot, we ask "How often do you talk with this farmer in an average month?". For this metric, we define $l_{ij} = 1$ if either i claimed to speak with j at least once per week or if j claimed to speak with i at least once per week. Respondents talked at least once per week in 18% of total matches.

We also use measures based on a roster of contacts for each respondent. We generated for each respondent a listing of all the individuals named by that respondent in a number of different contexts. The survey data includes people named in response to questions designed to record all 'significant' conversations about farming between individuals. In this case, we define $l_{ij} = 1$ if either i reports learning about farming from j, or if j reports learning about i. In principle, all pairs of respondents who have learned about fertilizer use from each other should have $l_{ij} = 1$. In fact, it is impossible to record the names of all individuals with whom respondents discuss farming, so enumerators were instructed to request the names of individuals with whom more than casual conversations took place. Two important issues arise as a consequence. First, different respondents interpret 'casual' differently, and second, important information might be transmitted during quite casual conversation.

We construct an additional metric using the roster of contacts that partly addresses this concern that information from casual contacts might be overlooked. Each roster of contacts also lists people who were hired by, borrowed from, lent or sold output to, or exchanged gifts, transacted land or jointly held assets with the respondent. We define an additional metric to examine the possibility that information about the outcome of experiments with fertilizer is transmitted when arranging any of these other transactions. For this metric $l_{ij} = 1$ if i appears anywhere in j's roster of contacts, or if j appears on i's roster.

In view of the potential drawbacks of the pairs measures yielding too small a subset of neighbors and the potential underreporting in the roster of contacts metrics we also construct predicted information neighborhoods. The metric we use is constructed from estimates of the probability of a positive response to the question "Have you ever gone to ___ for advice about your farm?" as a function of characteristics of both parties in the 7 randomly selected matches per respondent. Specifically, we estimate a probit model where the probability of a link is:

$$\Pr(l_{ij} = 1) = \Phi(m'_{ij}\alpha). \tag{5}$$

where Φ is a standard normal distribution function and m_{ij} is a set of indicators of plot proximity, gender, soil types, common religion, common clan, common parental background and migratory status. Estimates of α in equation (5) are presented in the Appendix. The most important variables for predicting links appear to be clan connections, gender, and physical proximity. We construct a predicted information neighborhood for farmer i as the union of all farmers j for whom the predicted $\Pr(l_{ij} = 1)$ is greater than a particular value.

We also experiment with constructing measures of second-order neighbors. Given any set of direct links l_{ij}, it is possible to define $l_{ij}^{II} = 1(\sum_{k \neq i,j} l_{ik} l_{kj} > 0)$ and the associated *second-tier* neighborhood $N_i^{II} = \{j | l_{ij}^{II} = 1\}$. Hence it is possible to examine the relationship between information availability and outcomes at different removes in the information network.

3.2 Economic and Agronomic Data

We focus on farmers' decisions about the use of chemical inputs. Many aspects of the pineapple growing technology are new to these farmers including planting materials, cultivation techniques, and marketing arrangements. Perhaps the most significant departure from traditional techniques is that a number of entirely new chemical inputs are used intensively in the cultivation of pineapple. There is agronomic evidence that pineapple yields are very responsive to these inputs and that their impact varies with local soil chemistry and weather patterns (Abutiate (1991); Purseglove (1972)). In informal interviews, individuals in the sample villages expressed conflicting views regarding the optimal timing, frequency and quantities of chemical inputs. There are official recommendations on fertilizer use available from the extension service of the Ministry of Agriculture, but these far exceed the levels of application used in these villages.[14] Uncertainty regarding the optimal application of chemical inputs certainly is not the only source of doubt regarding the technologies of pineapple production in these villages, but it provides a useful focus for our discussion of evidence regarding social learning.

The specific input decision we consider is the application of fertilizer per

[14]The recommendation is 400 Kg. of fertilizer/hectare, which is more than 10 times the mean fertilizer application observed in our sample. Only 4 of the 208 plantings we observed exceeded the recommended level of fertilizer application.

plant during the period from six weeks after planting through six months after planting. During this period, pineapples are extremely sensitive to nutrient availability (Bartholomew and Kadzimin (1977); Soler(1992)). Pineapple is not strongly seasonal because it can be chemically forced to flower and thus fruit at any time during the year in southern Ghana. Hence we observe pineapple being planted at each round in our survey data. For each observed 'planting', we calculate the per-plant value of fertilizer applied during the reference period after planting.[15] We also calculate the profits earned on these plantings, again on a per-plant basis.[16] We calculate profits by deducting the value of all inputs, including family labor valued at the relevant gender-specific wage from the value of output. Plot inputs and outputs were recorded at approximately six-week intervals over the two-year survey period.

Table 1 shows our estimates of the sensitivity of profit per plant to the application of fertilizer during this period. The point estimates indicate that profits per plant are maximized at a fertilizer input of approximately 18 cedis (the Ghanaian currency) per plant, which is about the ninety-fifth percentile of the distribution of fertilizer inputs over these plantings.[17] Profits seem to be increasing with other chemical applications on plots that are sloped or steeply sloped over the range of our data. These numbers, of course, should not be taken at face value as estimates of a production function, for they are plagued by the usual problems of simultaneity bias, which we make no attempt to address in this paper. Rather, they should be interpreted simply as evidence of the relevance of these particular measures of input intensity for the profitability of pineapple cultivation.

We also use detailed information about the plots themselves. Agronomic descriptions of each plot include its visible soil type, pH, and organic matter tests. Soil pH and organic matter content are key indicators of soil fertility

[15] Fertilizer prices were close to constant over our short sample time span and across villages. So reported values are essentially equivalent with amounts.

[16] Actual harvests are observed only for those plantings which occurred very early in our fieldwork. Plantings after round 5 were not harvested before the fieldwork ended. In the last round of the survey, respondents were asked for the price at which they could sell the crops that were currently standing on their plots. This is a normal transaction for crops such as cassava, in which plots full of partially grown plants are commonly sold to traders, who then hire labor to complete cultivation and harvesting of the plots. This kind of transaction is rare for pineapple, but respondents' familiarity with such sales of other crops made it easy for them to place values on their standing crops.

[17] The exchange rate ranged between approximately 1700 and 2300 cedis per $1 over the survey period.

in fallow farming like that practiced in the study area; these were measured at the start of the survey and analyzed at the Soil Science department of the University of Ghana.[18] In addition, all plots were mapped using global positioning system equipment. This procedure yields much more accurate measures of plot size and location than are available in most surveys in LDCs. It also allows us to control for spatial correlation in unobservables as a function of physical distances between plots. Moreover, it makes it possible for us to distinguish explicitly between the effects of information connections and those of geographical proximity.

The data also includes descriptions of a wide variety of transactions, and the roster of contacts information described above. Recorded contacts include learning interactions, credit and gift transactions, and labor market interactions. Data is recorded on the relationship and frequency of contact between the respondent and the contact, the residence and occupation of the contact, and the identification number of the contact if he/she is in the sample.

We are able to define neighborhoods based on the physical location of plots and reported financial transactions to contrast with those defined using the information metrics described above. We define the *financial* neighborhood of each farmer using a particular subset of the roster of contacts connections. In this case, $l_{ij} = 1$ if i or j report lending to, borrowing from, or exchanging gifts or holding assets in common with the other. *Geographical* neighborhoods of farmers are based on the proximity of their plots. Links are present, $l_{ij} = 1$, if the average distance between the plots cultivated by farmer i and farmer j is less than 500 meters.[19]

[18] See Goldstein and Udry (1999) for more detailed descriptions of the soil characteristics data.

[19] We have used a variety of different cutoffs to define the geographical link, including 1000 meters and 250 meters. We have also used the minimum distance between the plots cultivated by two farmers to define the link, with a cutoff of 250 meters. In no case were there substantive changes to the results reported in Section 5, with the exception that with very large cutoffs (over 1000 meters), virtually all farmers in the village are included in each other's 'geographical neighborhood,' and this variable becomes virtually collinear with the village $\times$ round effects used in some specifications.

3.3 Who Do Farmers Talk To and What Do They Know?

Contrary to conventional wisdom, it is certainly not the case that 'everyone knows everything' in these villages. Specific quantitative information about inputs and outputs is rarely shared: in less than two percent of 1878 random matches of a respondent to another farmer's recently harvested pineapple plot in the same village was the respondent able to provide an estimate of the quantity of pineapple harvested or fertilizer used.[20] Not only is information within the village not complete, but these numbers imply that the extent of information spillover is small because the villages are themselves small. The expected number of plots about which a farmer in these three villages would be able to provide quantitative information ranges from less than 1 in the smallest village to less than 7 in the largest village. Respondents report that they rely on conversation to learn about each others' farming activities; direct observation is a relatively unimportant source of information. Conversations about farming are more likely to focus on successful experiences than on failures; this observation (due to the work of Robert Afedoe in one of the sample villages) informs our regression specification.

Information links and geographical proximity are not strongly related within these villages. Figure 1 plots the geographical location and information neighborhood of each pineapple farmer in one of the survey villages. The geographical location of each farmer (as summarized by the average position of his pineapple plots) is indicated by the location of the vertices of the graph. Information links are indicated by the edges of the graph. In this graph, we have defined information links based on the farmers listed in each

[20]The survey was designed so that the responses of individual could be checked against the actual inputs and harvests reported by the farmer of the relevant plot. If the check is valid, the accuracy of the responses is astonishingly low. In less than 25 percent of the cases in which a respondent ventured to provide quantitative information was his/her report within 50 percent of the value reported by the farmer of the plot. Unfortunately, the questionnaire was not specific enough about the relevant period over which neighbors' inputs and outputs should be evaluated to ensure that this check is valid.

In addition, these figures may overstate the proportion of respondents who can provide quantitative information about other farmers' plots. After the first dozen (or so) interviews showed that very few respondents knew quantitative information on the plots of other farmers, we experimented with encouraging respondents to guess. This is a second potential explanation for the apparant poor acuracy of the reports of quantities. Because of these problems, we do not define information links between farmers on the basis of this variable.

respondent's roster of contacts. Similar graphs can be drawn for the other villages, using alternative definitions of information links, and this conclusion remains unchanged: neighborhoods defined by geographical proximity and by information connections are not coincident. Figure 2 shows estimates of the density of distances between farmers who have and do not have information links. Farmers whose plots are closer together are more likely to have discussed farming, but there is a large overlap in the two distributions.

Thus there is evidence that 'collective experimentation' in which the outcome of every experiment with a new technology by a farmer in a village is revealed, as in Example 1, is not a suitable benchmark for these villages. There appears to be both incomplete observability of one's neighbors' experiments with the new technology (akin to Example 2) *and* a nontrivial neighborhood structure as in Example 3.

4 Empirical Model of Updating

There is a great deal of variation in the level of input application, both across the sample and over time by individuals. For example, Figure 3 portrays this variation for farmers in one of the sample villages. Line segments join successive plantings of pineapple by a particular farmer, so it can be seen that some farmers increase their use of fertilizer, while others reduce it. Averaged over farmers, the standard deviation over time of fertilizer use is 3.2 (mean fertilizer use is 3.0 cedis/plant). There are 34 instances when a farmer simultaneously planted more than one plot with pineapple. Averaged over these instances, the standard deviation of fertilizer use is 1.2. As expected, there is a much larger variation in input use by individual farmers over time than across plots at a point in time.

To investigate the importance of learning, we estimate whether farmers change their fertilizer use to align with that of their information neighbors who were successful in similar circumstances. The simplest version of this comparison of outcomes and actions is illustrated in Table 2. Panel A summarizes the outcomes of the experiments with fertilizer over the survey period by our respondents. Entries are percentages of plantings with all combinations of higher and lower than median profits and zero and positive fertilizer use. Panel B limits attention to pineapple plantings by farmers who are in the information neighborhood of respondents who increased fertilizer use over the survey period, and Panel C limits attention to pineapple plantings by farmers who are in the information neighborhoods of respondents who did *not* increase fertilizer use over the survey period.[21] If we compare Panels B and C, we see that a larger percentage of plantings with high fertilizer use were associated with low profits in the experiments by farmers in the neighborhoods of respondents who did not increase fertilizer use compared with those in the neighborhoods of respondents who did increase fertilizer use. Thus increases in fertilizer use by a farmer are associated with the outcomes of experiments with fertilizer by other farmers known to that farmer. This pattern is consistent with the hypothesis that social learning plays a role in innovation in these villages. The remainder of our empirical work is devoted to examining whether this basic relationship holds when appropriate conditioning information is used to distinguish this hypothesis from alternatives that might give rise to the same cross-tabulation.

Regression Specification

For simplicity, we present the notation as though each farmer has only one pineapple plot. Defining notation, let $f_{i,t}$ be the value of fertilizer used per new plant applied by farmer i for the pineapples planted in round t. We let the characteristics of i and his plot be contained in a vector $\omega_{i,t}$. Farmer i's profits per plant for the round t planting are denoted $\pi_{i,t}$ and the set of farmer $i's$ neighbors is given by N_i which, with some abuse of notation, is also used to denote the number of neighbors.

In order to quantify relative success of an individual we use a simple function of whether his realized profits fall above a 'typical' profit level π^*. This corresponds to the limited observability of Example 2. In principle, this typical profit level could be allowed to be a time-varying, individual- or plot-specific forecast of profits. However, because of our data's short time span, we use median profits in our sample to define π^*. We note that we could also construct measures of failure and test whether farmers avoided the actions of their unsuccessful neighbors. We focus on measures of success because farmers are more likely to announce successes than failures to their neighbors.

Using the notation $\Delta f_{i,t}$ for the first difference of inputs, $(f_{i,t} - f_{i,t-1})$,

[21] In both cases, the information neighborhood is based on positive responses to the question: "Have you ever gone to ___ for advice about your farm?". Similar results are obtained when using the alternative definitions of information neighborhood.

our basic updating equation is:

$$\Delta f_{i,t} = \omega'_{i,t}\alpha + \beta \frac{1}{N_i} \sum_{j \in N_i} 1\left(\pi_{j,t-1} - \pi^*\right)\left[f_{j,t-1} - f_{i,t-1}\right] + \delta A_{i,t} + \varepsilon_{i,t} \tag{6}$$

where $1(\cdot)$ is an indicator function equal to one if its argument is positive and zero otherwise. The first term in this specification reflects the influence of individual and plot specific components on the change in fertilizer used on new plantings. The second term:

$$\frac{1}{N_i} \sum_{j \in N_i} 1\left(\pi_{j,t-1} - \pi^*\right)\left[f_{j,t-1} - f_{i,t-1}\right]. \tag{7}$$

is meant to reflect social learning contributions. Any of the N_i farmers in the information neighborhood of farmer i affects farmer $i's$ input change if he is successful so that $1\left(\pi_{j,t-1} - \pi^*\right)$ is one. Each successful neighbor adds to this index a term $\left[f_{j,t-1} - f_{i,t-1}\right]$ that shifts farmer $i's$ inputs toward those of his successful neighbor. This term, therefore, reflects the main conclusion of Section 2 that a farmer updates his beliefs regarding the optimal use of fertilizer in the direction of successful experiments with fertilizer by other farmers in his information neighborhood. When implemented in the following section, term (7) will include all previous period plantings by farmers in N_i, not the one per farmer assumed here for notational simplicity. The third term $A_{i,t}$ consists of variables that would help forecast changes in $f_{i,t}$ under various alternative explanations, as discussed below. The final, unobservable $\varepsilon_{i,t}$ term is assumed to be uncorrelated with the regressors but allowed to be correlated across plots as a general function of their physical distance using the spatial GMM estimation strategy in Conley (1999).[22]

There are several important limitations of specification (6). The model of learning in Section 2 implies a nonstationarity in updating changes that

[22] Our estimates use limiting results for cross section estimation with spatial dependence characterized by physical distance. Serial dependence is allowed for only by use of time (round) dummies. Spatial standard errors are calculated using the estimator in Conley (1999) with a weighting function that is the product of one kernel in each dimension (North -South, East-West). In each dimension, the kernel starts at one and decreases linearly until it is zero at a distance of 1.5 km and remains at zero for larger distances. This estimator is analogous to a Newey-West (1987) time series covariance estimator and allows general correlation patterns up to the cutoff distances. The inferences reported below are robust to cutoff distances between 1 km and 2 km.

is not captured by this specification. In particular, as agents learn about the target input distribution the variance in their input changes should go to zero if the target input distribution is fixed. This effect is not captured in model (6), so (6) must be interpreted as a linear approximation to the updating rule. In addition, farmers might learn from their own experience as well as from the experience of other farmers in their information neighborhood. But outcomes of lagged experiments by farmer i are not included in (6). The short time span of our data precludes any attempt to identify either the impact on the rate of change of updating associated with the declining variance of farmer posterior beliefs or the impact of learning-by-doing by an individual farmer over time.

Another limitation of (6) is that the learning term (7) is essentially a function neighbors' average behavior so it does not capture the increased precision of signals coming from a larger number of neighbors. This is more of a concern for our roster of contacts metrics than our random sample metrics as the former results in a more disparate number of neighbors across farmers.

In addition, there several alternate explanations that would suggest that significant β estimates might be caused by omitted variable bias because information neighbors share common access to credit arrangements, or correlations caused by weather or soil variation. We are able to address many of these concerns by modifications of the basic specification (6) that vary the definition of $A_{i,t}$.

Correlation due to common weather shocks and unobserved soil characteristics are likely to result in successful input changes being related in geographically close areas. To investigate whether there is a spurious result arising from the fact that geographic proximity is related to information distance, in some specifications we define $A_{i,t}$ as a measure of geographic neighbors' input usage. This is constructed as in (7) with geographic rather than information neighbors. This is also a plausible check of the hypothesis that omitted own past history is the cause of significant β estimates, because geographical neighbors' current input usage will tend to be correlated with own past inputs insofar as inputs respond to shock histories that have significant geographic correlations (such as weather). We also construct similar regressors to reflect input usage by those within credit and gift networks– 'financial neighbors'. Finally, in all our specifications, our inferences allow for general spatial correlation in unobservables across farmers.

The aim of these regressions is to determine whether state variables that contain information about the profitability of fertilizer use are good predictors for observed actions. The parameters of equation (6) are not parameters of a structural model and are best interpreted as those of a linear prediction of innovations in behavior. As such, measurement error in definitions of N_i is not a major concern, because this would not lead to incorrect inference about partial correlations. Endogeneity of neighborhoods that arises from individuals choosing neighbors who would be good to learn from is also not a concern. Endogeneity of neighborhoods arising from unobservables (like wealth were it an omitted variable) that influence neighbor choices and profitability as well as innovations are, of course, an important concern as these unobservable effects would result in the same correlations as learning from high-profit neighbors.

5 Estimation Results

5.1 Do Experimental Outcomes in an Information Neighborhood Affect Cultivation Decisions?

Tables 3 through 6 present the results of estimating equation (6) and variations upon it. As described above, the dependent variable in these regressions is the change in fertilizer use by a farmer. The crucial independent variable is that labeled 'successful experiment in information neighborhood.' This is defined by equation (7) using N_i derived from information links based on positive responses to the the question: "Have you ever gone to ___ for advice about your farm?". The estimated coefficients on this variable are approximately one across all specifications. To interpret this magnitude, suppose that in period $t-1$ every farmer in the information neighborhood of farmer i used one unit more fertilizer than did farmer i. If two thirds of those farmers achieved higher than median profits, then the coefficient of approximately unity implies that the expected increase in farmer i's fertilizer use between $t-1$ and t is $2/3$.

Our baseline estimates are presented in Table 3. Column A is similar to the cross-tabulation presented in Table 2. Column B adds a set of fixed characteristics of farmer i : wealth and clan, religion, and village indicators.[23]

[23] We note that village-level dummies and standard errors are, strictly speaking, not consistently estimated with the spatial dependence model of Conley (1999) that we use. This is because limits are taken as the region of observation grows, analogous to growth

Column C adds a further set of characteristics of the plot: sand or loam soil types (clay is the omitted category), pH, and organic matter content. Data on soil type and the results of soil chemistry tests are not available for all plots, so the sample size falls in this column. The correlation between successful experiments in the farmer's information neighborhood and the change in the farmer's fertilizer use is stable across these specification changes. The revelation of information about the usefulness of fertilizer in the information neighborhood of a farmer is significantly correlated to innovations in that farmer's use of fertilizer.[24]

In column D of Table 3, we examine the hypothesis that the outcomes of experiments in the 2nd tier information neighborhood of a cultivator are correlated with changes in fertilizer use. The coefficient is positive and smaller than that associated with the 1st tier information neighborhood, but we cannot reject the hypothesis that there is no effect of this indirect information on cultivation.

Our primary concern regarding these estimates is the identification problem raised in the introduction. Could it be that the profitability of fertilizer use is increasing particularly rapidly amongst a set of farmers who share some characteristic omitted from the regressions estimated in Table 3? If this is the case, the correlation between successful experiments with fertilizer

in the span of a time series. The time series counterpart to a village dummy would be a dummy for a fixed interval, which could not be consistently estimated with a typical time series because the number of observations in, say, the 1970s would be fixed as the sample length grows. Nevertheless, if the spatial dependence is over a short distance relative to the size of a village the estimates of the village dummies and their standard errors will still provide useful approximations.

[24] $f_{i,t-1}$ appears on both sides of equation (6), raising the possibility that its measurement error may induce an important bias in our estimate of β. However, we believe that our field research strategy based on frequent visits throughout the survey period resulted in high quality data on fertilizer inputs in particular. Bias due to classical measurement error in $f_{i,t-1}$ would be approximately towards the overall proportion of profitable farmers in neighborhoods rather than to the vicinity of 1, where our estimates lie. For example in the information neighborhood as defined in Table 3, the average proportion of successful experiments is .3; when the information is defined based on farmers' rosters of people with whom they have interacted, the proportion is .4. As a further check that measurement error is not the cause of our β estimates being consistently near one, we experimented with artifically adding measurment error to our $f_{i,t-1}$ observations and rerunning regressions. In all cases, using the noisy version of $f_{i,t-1}$ results in substantially lower estimated coefficients.

by people known to a farmer and increases in fertilizer use by that farmer may reflect not social learning but that unobserved characteristic. Many such characteristics would be spatially correlated according to geographic proximity, just like information links, hence the difficulty of distinguishing the effects of social learning from such factors.

Table 4 presents estimates of variations of specification (6) that include geographic and financial neighborhood regressors. In column A, we show that there is indeed a strong correlation between successful experiments with fertilizer in a geographic neighborhood of a farmer and innovations in fertilizer use by that farmer - almost as strong as that which we found for successful experiments in the information neighborhood of the farmer.[25] That correlation, however, appears to be entirely an artifact of the spatial (geographic) correlation of information links: in column B, it is shown that successful experiments in the geographic neighborhood of a farmer are not significantly related to innovations in fertilizer use, conditional on the proportion of successful experiments in the farmer's information neighborhood. Moreover, the magnitude of the estimated effect of successful experiments in the information neighborhood is virtually unaffected by the inclusion of the geographic neighborhood variable.

Robustness of our estimated information neighborhood effects to the inclusion of geographic neighborhood variables is also a positive sign regarding correlations resulting solely from the lack of individual histories. Reactions to spatially (geographically) correlated weather shocks and unobserved soil characteristics are likely to cause geographic neighbors' variables to be correlated with own past histories. If so, the fact that the information neighborhood effect is stable when the geographic neighborhood variable is included in the specification is evidence of minimal bias resulting from the omission of the farmer's history of fertilizer innovations from the regression.

It is likely that information links are correlated not only with geographic proximity but with other social and economic ties. Given the relatively high working capital demands of pineapple production, it is possible that cultivation decisions are influenced by short-term fluctuations in the financial environment. Could the innovations in fertilizer use be a consequence of financial shocks transmitted through social networks that are largely parallel to the information networks we have examined? We examine this potential confounding factor in column C of Table 4. Successful experiments in the

[25] The geographical neighborhood is defined as described in section 3.

financial neighborhood of a farmer are not significantly positively related to innovations in fertilizer use by that farmer; indeed, the correlation is significantly negative.[26] Inclusion of the financial neighborhood variable has almost no effect on the magnitude or statistical significance of the estimate of the effect of experiments in the information neighborhood.

In Table 5, we extend the specifications to allow for arbitrary village and round effects on changes in fertilizer use. In column A, we see that the inclusion of round effects has almost no effect on the size or significance of the estimate of the effect of successful experiments in the information neighborhood of a farmer on that farmer's innovations in fertilizer use. In column B, we include a set of village-round interactions.[27] The coefficients of these interactions could reflect changes in prices, weather shocks or other unobserved village-level events that effect the expected profitability of fertilizer use. This is the regression equivalent of the conventional differences-in-differences estimator: do farmers in whose information neighborhood a successful experiment with fertilizer occurred last period increase their fertilizer use by more than other farmers in the same village in the same period? The answer is yes: there are significant village-round effects (growth in fertilizer use seems to have been particularly slow in the middle rounds in village 3), but conditional on those effects, the estimated coefficient on successful experiments in the information neighborhood remains almost unchanged in magnitude and statistical significance from other specifications.

In column C, we include individual covariates along with the village-round interactions, with no important change in the results. In column D, we show that the main effect of successful experiments in the information neighborhood remains robust to the inclusion of the analogous measure of successful experiments in the financial neighborhood in this new specification.

[26] The significant negative relationship that we find here between successful experiments with fertilizer in the financial neighborhood of a cultivator and that cultivator's innovations in fertilizer use is not robust to alternative specifications. In particular, it is not robust to the inclusion of round effects (see table 5, column D). The financial neighborhood is defined as described in section 3.

[27] Only 16 plantings occurred in village 1, so a separate set of village 1 $\times$ round interactions is not included. Even with this consolidation, the round by village cell sizes in this regression are small and thus standard errors and inference regarding the village $\times$ round interactions should be interpreted with considerable caution.

5.2 Alternate Measures of the Information Neighborhood

As described in Section 3, we have available a variety of alternative measures that capture different dimensions of the relationship we refer to as an "information link." In Section 5.1 we defined links based on positive responses to the question, "Have you ever gone to ___ for advice about your farm?". In Table 6, we present results based on six alternative definitions of information links. In each case, only the coefficient (and standard error) of the "successful experiment in the information neighborhood" is reported.[28] For each definition, we report the results of two specifications of the updating process. The first specification (analogous to that reported in Column C of Table 4) includes as additional covariates a set of farmer characteristics, and measures of successful experiments in the geographical and financial neighborhoods of the respondent. The second specification (analogous to that reported in Column C of Table 5) includes in addition a set of round dummy variables and interactions between the round dummy variables and the village 3 indicator. The first three rows of the table report results based on the random matching of respondents with recently-harvested pineapple plots cultivated by other farmers in the village. The first row uses the most liberal definition of an information link, based simply on whether the respondent knows the other farmer. The second row narrows the definition, limiting links to cases in which respondents know about the particular randomly-matched pineapple plot. The third row further limits the definition to situations in which the respondent knows about the plot and talks with the other farmer at least once per week. When neighborhoods are defined based on any of these definitions of the information link, there is a strong and significant positive correlation between successful experiments with fertilizer use in a farmer's information neighborhood and that farmer's subsequent innovation in the use of fertilizer.

The following two rows report results with information neighborhoods defined by listings of individuals with whom respondents had dealings. In the first of these, information links are defined based on individuals named by respondents when asked a series of questions about who taught them to farm and to whom they have given farming advice. Based on this definition of

[28] The full results are available from the authors. The correlation between the neighborhood information scores based on these different definitions of information links is reported in Appendix Table 3.

information links, successful experiments with fertilizer in a farmer's information neighborhood are strongly and significantly correlated with innovations in fertilizer use by the farmer. In the fifth row, we report results using a broader definition of an information link. In this case, we define a links based on not only explicit 'learning interactions', but also on any market exchange over a ten month period. The motivation for the use of this definition is that information about the outcome of experiments might be transmitted in a very informal manner, in the context of interactions motivated by other purposes. Recall that in all these specifications, the index of successful innovations in the 'financial neighborhood' of the farmer is included as a covariate. The 'financial neighborhood' is a subset of the information neighborhood based on this definition of information links. Again, we see that there is a strong and significant correlation between innovations in fertilizer use by a farmer and successful experiments with fertilizer by other individuals with whom that farmer has an information link.[29]

The final row of Table 6 reports results of estimating equation (6) when we define the information neighborhood based on predicted links using (5). Each respondent's neighborhood includes all the sample farmers with whom his predicted link probability is greater than 5%. Again, we see that there is a strong and significant correlation between innovations in fertilizer use by a farmer and successful experiments with fertilizer by other individuals with whom that farmer is predicted to have an information link.

Thus far, we have examined the hypothesis that a farmer adjusts his fertilizer use towards that of others in his information neighborhood who have achieved high profits. An alternative hypothesis is mimicry: do farmers copy the behavior of the people they talk with, without judging the success of that behavior? To test this alternative hypothesis, we modified our updating model to include a term designed to capture mimicry. We constructed a measure of fertilizer use by *all* the farmers in the information neighborhood of the cultivator, not just those with high profits: $A_{i,t} \equiv \frac{1}{N_i} \sum_{j \in N_i} [f_{j,t-1} - f_{i,t-1}]$. Unfortunately, our sample does not contain enough variation to separately identify the effect of this $A_{i,t}$ term from our learning term (7). The precision of estimates was generally poor and inferences about coefficients were not robust across neighborhood metrics nor across specifications with different

[29] Conditional on succesful innovations in the information neighborhood, there is no significant correlation between successful innovations in the financial neighborhood and fertilizer innovations by the farmer.

conditioning variables, therefore we do not report them.

6 Conclusion

This paper presents evidence that social learning is important in the diffusion of knowledge regarding pineapple cultivation in Ghana. We take advantage of data that combines agronomic and conventional economic information with details regarding relationships between farmers to address the challenge of identifying learning effects in an economy undergoing rapid technological change. We trace the effect of a farmer's successful experiment with fertilizer on the innovations in fertilizer use by other cultivators with whom the experimenter shares information. We find that farmer A increases (decreases) his fertilizer use after someone with whom he shares information achieves higher than median profits when using more (less) fertilizer than farmer A. Conditional on this effect, we find no evidence that farmer A changes his fertilizer use after a farmer who cultivates a nearby plot achieves high profits when using more or less fertilizer than farmer A. Nor is there evidence that farmer A changes his fertilizer use after a farmer with whom he has financial dealings achieves high profits when using more or less fertilizer than farmer A. All of these conclusions are robust across alternative definitions of 'sharing information,' to a variety of definitions of 'nearby plot,' and for specifications that condition on a range of individual and plot characteristics and on arbitrary village and round effects. This finding conforms with the robust conclusion of the examples of social learning described in section 2: successful experiments with input use attract imitation. Moreover, our empirical results establish that this imitation occurs within social groups defined by information flow, and not by geographical proximity or financial interconnections.

It is profitable for farmers to adjust their fertilizer use towards that of their successful neighbors. Table 7 reports the results of estimating an equation akin to (6) when $\Delta f_{i,t}$ is replaced on the left hand side with $\Delta\pi_{i,t}$. A one standard deviation increase in the index of successful experiments in a farmer's information neighborhood is associated with an increase in that farmer's profits (over that farmer's previous planting) of between 25 and 35 percent of mean profits per plant.[30]

[30] The standard deviation of the index of successful experiments in farmers' information neighborhoods ($\frac{1}{N_i}\sum_{j\in N_i} 1\left(\pi_{j,t-1}-\pi^*\right)\left[f_{j,t-1}-f_{i,t-1}\right]$) is 6.32, and the mean profit per

There is evidence that social learning plays a role in the cultivation decisions of these farmers. Information, therefore, has value in these villages, as do the network connections through which that information flows. This raises the possibility that farmers consider the consequences for the availability of information when forming the connections that underlie their information neighborhoods. If so, measurement of the extent of social learning is not sufficient for adequate evaluation of policy regarding the diffusion of technology. It is necessary, in addition, to understand the endogenous process of information network formation. For example, consider the impact of a subsidy offered to one farmer in a village that induces him to use an optimal large amount of fertilizer and (with high probability) get high profits. The speed with which this information spreads, and hence the value of the subsidy, depends upon the choices of the subsidized farmer and others in the village to make and maintain information linkages. These choices may depend upon the value of the information to each farmer and upon the costs of information links, which may depend upon a rich array of characteristics of the farmers and the social structure of the village. In some contexts, differing religions may be an effective barrier to communication. In others, gender, wealth or family ties may be the most salient determinants of the shape of the information network.

The next step in this research program is to model the choices of farmers regarding the formation of information links in these villages. A large literature examines network efficiency, with the goal of characterizing the network configuration that maximizes a value function (Bolton and Dewatripont (1994); Hendricks, Piccione et al. (1995); Economides (1996)). This is appropriate for a planner (such as a telecommunications monopoly) but not for the decentralized process that governs the formation of a social network like those in the sample villages. Accordingly, we focus on the incentives of the individuals who build the links that define the network ((Coleman (1966); Granovetter (1973); Granovetter (1992); Fafchamps (1999); Fafchamps and Minten (1999)). Bala and Goyal (1999), Jackson and Wolinsky (1996) and Montgomery (1996) provide alternative theoretical schema that can underpin an empirical analysis of link formation, but we know of no empirical work in

plant is 253 cedis. An alternative perspective on the magnitude of these estimates is to consider the case in which a farmer's neighbors all use 1 cedi more fertilizer per plant than the farmer, and that two-thirds of these neighbors are successful. The estimates imply that this will be associated with an increase in profits of approximately 7 to 9 cedis per plant.

economics that examines the formation of decentralized networks. It is possible that this process can be identified in these data, because we observe the formation of links over time, as well as many details about the relationship between individuals and much of the salient agronomic information.

References

[1] Abutiate, W. S. (1991). Guide to the Commercial Production of Pineapple, Papaya and Mango in Ghana.

[2] Aghion, P. and P. Howitt (1998) *Endogenous Growth Theory.* Cambridge, MIT Press.

[3] Bala, V. and S. Goyal (1999). "A Non-Cooperative Theory of Network Formation." *Econometrica* forthcoming.

[4] Bartholomew, D. and S. Kadzimin (1977). Pineapple. Ecophysiology of Tropical Crops. P. d. T. Alvim and T. T. Kozlowski. New York, Academic Press: 113-156.

[5] Besley, T. and A. Case (1994). "Diffusion as a Learning Process:Evidence from HYV Cotton." Working Paper.

[6] Bolton, P. and M. Dewatripont (1994). "The Firm as a Communication Network." *Quarterly Journal of Economics* 109(4): 809-840.

[7] Bindlish, V. and R. E. Evenson (1997). "The Impact of T&V Extension in Africa: The Experience of Kenya and Burkina Faso." World Bank Research Observer 12(2): 183-201.

[8] Birkhaeuser, D., R. E. Evenson, et al. (1991). "The Economic Impact of Agricultural Extension: A Review." Economic Development and Cultural Change 39(3): 607-650.

[9] Brock, W. and S. Durlauf (1999) "Interactions-Based Models," Forthcoming in *Handbook of Econometrics,* J. Heckmand and E. Leamer eds.

[10] Conley, T. (1999) "GMM Estimation with Cross Sectional Dependence" *Journal of Econometrics* Vol 92(1) p.1-45.

[11] Coleman, J. E. Katz and H. Menzel (1957) "The Diffusion of an Innovation Among Physicians," *Sociometry* 20: 253-270.

[12] Coleman, J. (1966). *Medical Innovation: A Diffusion Study.* New York, Bobbs-Merrill.

[13] Economides, N. (1996). "The Economics of Networks." *International Journal of Industrial Organization* 14(6): 673-99.

[14] Evenson, R. and L. Westphal (1995). Technological Change and Technology Strategy. Handbook of Development Economics. J. Behrman and T. N. Srinivasan. Amsterdam, North-Holland. 3A: 2209-2300.

[15] Fafchamps, M. (1999). Ethnicity and Credit in African Manufacturing. Manuscript: Oxford.

[16] Fafchamps, M. and B. Minten (1999). "Relationships and traders in Madagascar." *Journal of Development Studies* 35(6): 1-35.

[17] Feder, G., R. Just, et al. (1985). "Adoption of Agricultural Innovations in Developing Countries: A Survey." Economic Development and Cultural Change 33(2): 255-298.

[18] Foster, A. and M. Rosenzweig [1995] "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture" *Journal of Political Economy* Vol. 103(6) p 1176-1209.

[19] Goldstein, M. and C. Udry (1999). Gender and Land Resource Mangement in Southern Ghana.Manuscript: Yale University.

[20] Granovetter, M. (1973). "Strength of Weak Ties." American Journal of Sociology 78(6): 1360-1380.

[21] Granovetter, M. (1992). "Economic Institutions As Social Constructions - a Framework For Analysis." *Acta Sociologica* 35(1): 3-11.

[22] Griliches, Z. (1957). "Hybrid Corn: An Exploration in the Economics of Technological Change." Econometrica 25(4): 501-522.

[23] Hendricks, K., M. Piccione, et al. (1995). "The Economics of Hubs: The Monopoly Case." *Review of Economic Studies* 62(1): 83-101.

[24] Jackson, M. O. and A. Wolinsky (1996). "A Strategic Model of Social and Economic Networks." *Journal of Economic Theory* 71(1): 44-74.

[25] Jovanovic, B. and Y. Nyarko (1994). The Bayesian Foundations of Learning by Doing. NYU, Dept. of Economics.

[26] Kapteyn, A. (2000) “Saving and Reference Groups” Tilburg University CentER Working Paper.

[27] Lucas, R. E. Jr. (1988) “On the Mechanics of Economic Development.” *Journal of Monetary Economics.* 22:3-42.

[28] Manski, C. F. (1993) “Identification of Social Endogenous Social Effects: The Reflection Problem” *Review of Economic Studies* 60: 531-542.

[29] Manski, C. F. (1997) *Identification Problems in the Social Sciences* Harvard University Press, Cambridge.

[30] Moffitt, R. (1999) “Policy Interventions, Low-Level Equilibria, and Social Interactions” Johns Hopkins Working Paper.

[31] Montgomery, James (1996) “The Structure of Social Exchange Networks: A Game-Theoretic Reformulation of Blau’s Model.” *Sociological Methodology* 26:193-225.

[32] Morris, S. (1995). “The Common Prior Assumption in Economic Theory.” *Economics and Philosophy* 11:227-253.

[33] Munshi, K. and J. Myaux (1998) “Social Effects in the Demographic Transition: Evidence from Matlab, Bangladesh” Working Paper Univeristy of Pennsylvania.

[34] Newey, W.K. and West, K.D. (1987). “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica* 55:703-708.

[35] Obeng, I. S. (1994). Effects of Domestic Policies on Production and Export of Non-Traditional Agricultural Commodities: A Case Study of Fresh Pineapples in Ghana, M.Phil Thesis, Department of Agricultural Economy and Farm Management, University of Ghana, Legon.

[36] Prescott, E. (1972). “The Multi-Period Control Problem Under Uncertainty.” *Econometrica* 40(6): 1043-1058.

[37] Purseglove, J.W. (1972) *Tropical Crops: Monocotyledons I.* London: Longman.

[38] Rogers, E. (1995) Diffusion of Innovations, 4th Edition. New York: Simon & Schuster.

[39] Romer, Paul (1990) "Endogenous Technical Change." *Journal of Political Economy.98:5 pt.2 S71.*

[40] Soler, Alain (1992) *Pineapple: Quality Criteria*. CIRAD-COLEACP-OCAB, Paris.

[41] Wilson, R. (1975). "Informational Economies of Scale." *Bell Journal of Economics*: 184-195.

[42] Woittiez, I. and Kapteyn A. (1998) "Social interactions and habit formation in a model of female labour supply" *J. Public Economics* (70)2 p. 185-205.

[43] Yamauchi K, Futoshi (2000) "Social Learning and Investment in Human Capital: Evidence from Schooling Investments in India" Working Paper Kyoto University.

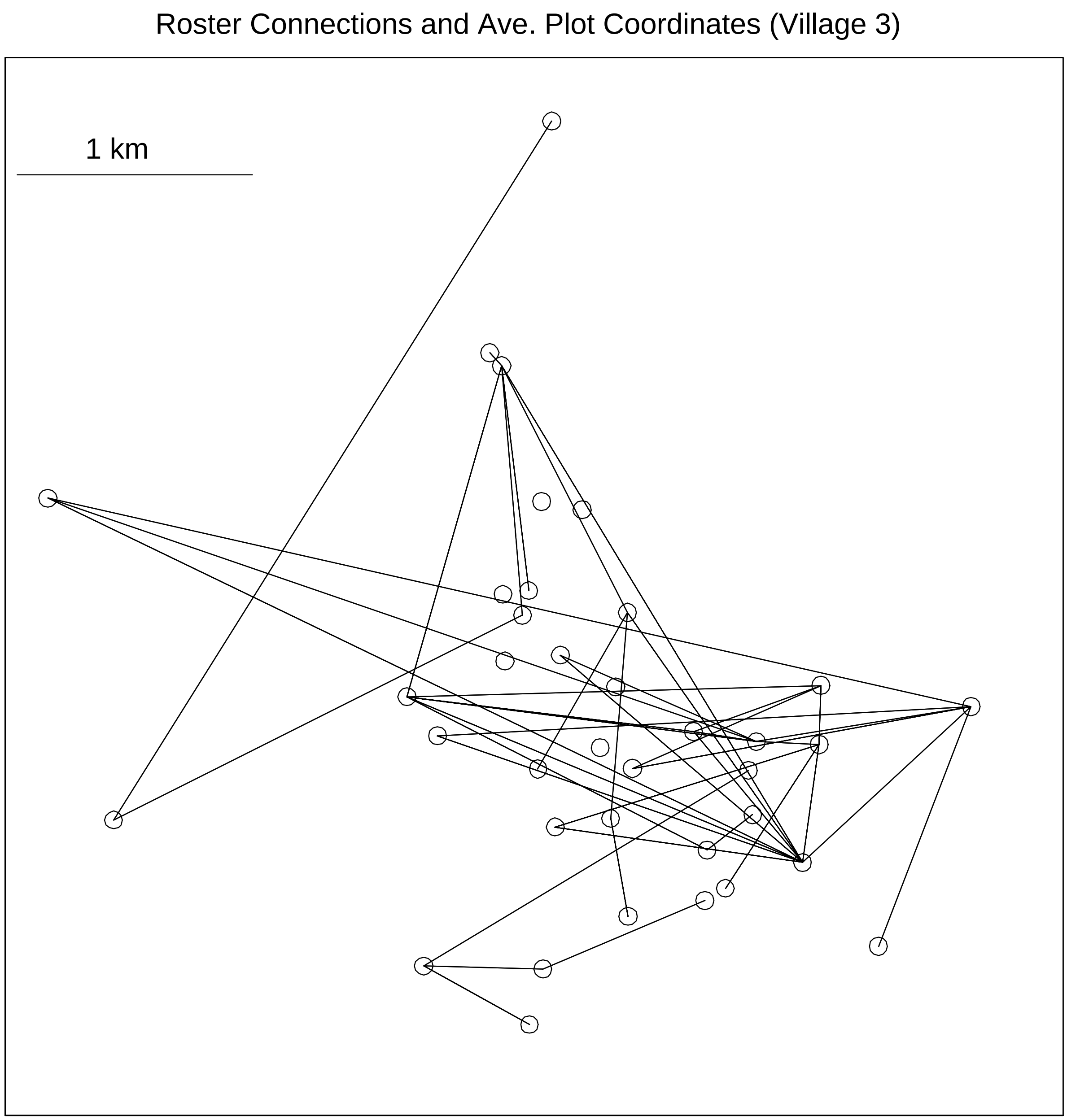

Figure 1. The circles represent the geographic center of plots for each farmer in Village 3. The lines connect farmers listed on each other's roster of contacts as having had significant converstations about farming.

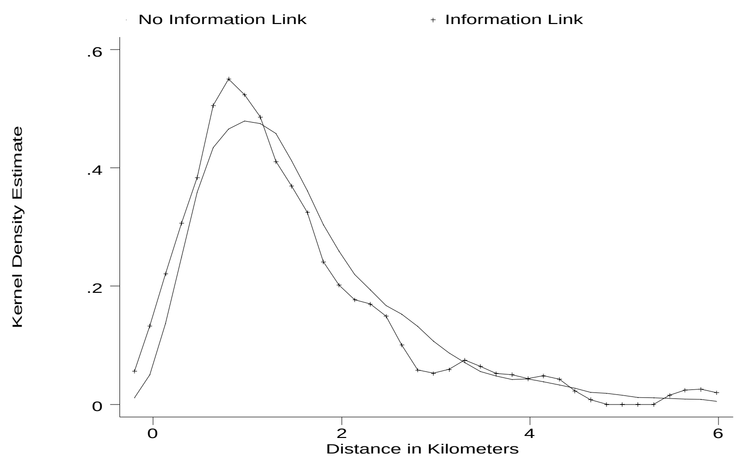

Figure 2. Kernel density estimates for the distribution of distances between all farmers and those listed on the roster of contacts as having had significant conversations about farming. An Epanechnikov kernel with a bandwidth of .2 was used.

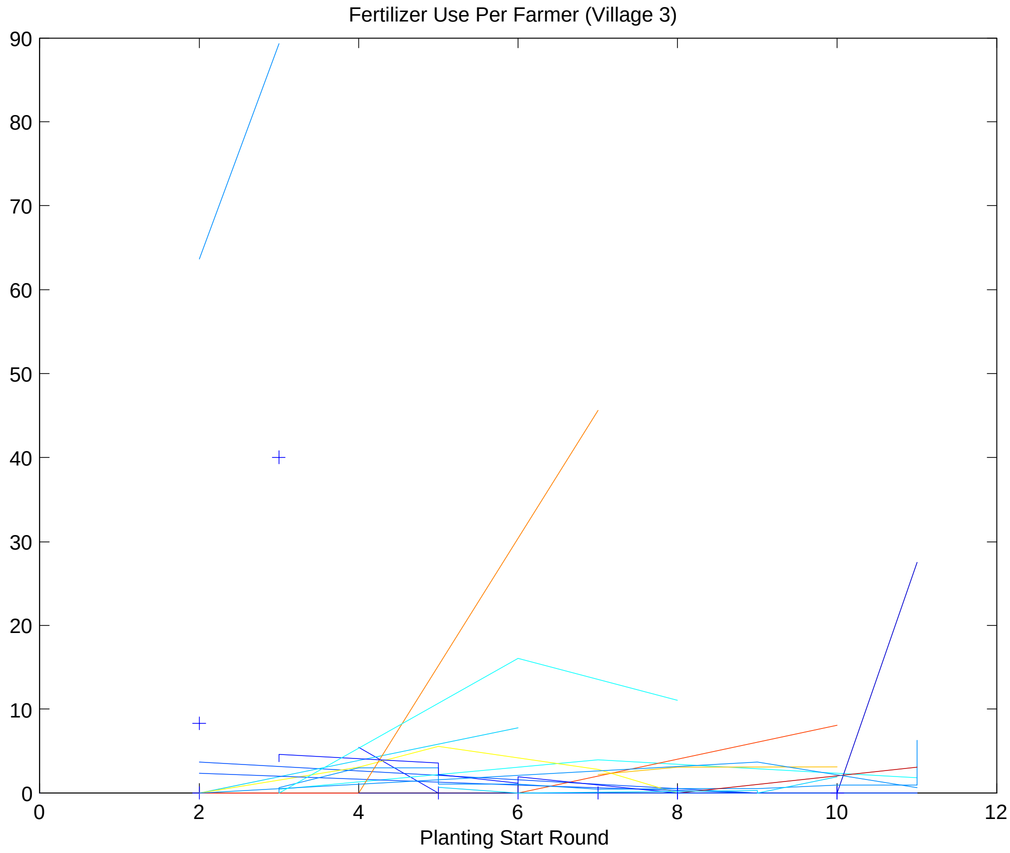

Figure 3. This figure plots fertilizer use versus starting round for all plantings in Village 3. Line segments join successive plantings of pineapple by a particular farmer and plusses indicate farmers with a single planting.

Table 1: Relationship between Profit/Plant and Chemical Inputs

LAD regression Dependent Variable Profits/Plant (cedis)

Regressor	Coefficient	Std. Error
Constant	80.30	167.29
Slope	-29.63	38.48
Steep Slope	-140.06	67.01
Fertilizer/plant	14.95	5.17
(Fertilizer/Plant)^2	-0.41	0.10
Other Chemicals/Plant	0.60	2.73
Other Chemicals*Slope	14.84	3.21
Other Chemicals*Steep Slope	95.25	21.21
pH	53.96	28.40
% Organic Matter	-90.10	18.65

n=164, Median Profits/Plant = 125 cedis

*Standard error calculation assumes independence across observations.

Table 2: Fertilizer Use and Profits, by Information Neighborhood

Panel A: Fertilizer Use and Profits in the Sample of Plantings (208 plantings, cell percentages)

	Fertilizer Use	
Profits	None	Positive
Lower than Median	37	12
Higher then Median	31	21

Panel B: Fertilizer Use and Profits in the Sample of Plantings of People Who Are
In The Information Neighborhoods of Those Who Increased Fertilizer Use (66 plantings)

	Fertilzer Use	
Profits	None	Positive
Lower than Median	35	10
Higher then Median	35	20

Panel B: Fertilizer Use and Profits in the Sample of Plantings of People Who Are
in the Information Neighborhoods of Those Who Did Not Increase Fertilizer Use (142 plantings)

	Fertilizer Use	
Profits	None	Positive
Lower than Median	25	22
Higher then Median	34	19

Table 3: Innovations in Fertilizer Use

Dependent Variable: Change in Fertilizer Use Per Plant

Regressor	A		B		C		D	
	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.
Successful Experiment in Neighborhood	0.993	0.188	0.999	0.149	0.913	0.173	0.657	0.285
Successful Exp. in 2nd-tier Neighborhood							0.083	0.075
Constant	-0.162	0.674	0.509	1.395	-21.350	13.085	0.935	1.326
Village 1			-4.288	1.761	-7.512	2.883	-8.783	4.617
Village 2			1.000	1.334	1.897	2.173	0.161	1.425
Wealth			-0.419	0.436	-0.381	0.417	-0.288	0.490
Clan 1			-0.021	1.287	-0.719	2.031	-1.092	1.548
Clan 2			0.475	1.295	1.491	1.554	-0.417	1.630
Religion 1			0.872	1.654	2.289	1.935	0.814	1.677
Sand					2.404	1.322		
Loam					1.077	2.446		
Organic Matter					0.752	0.737		
pH					3.191	2.546		
	n=107		n=107		n=87		n=107	

Spatial standard errors calculated as described in Footnote 22. Successful innovations in neighborhood variables as defined in equation (7) with neighborhoods defined using responses to: "Have you ever gone to ___ for advice about your farm?"

Table 4: Innovations in Fertilizer Use, with Information, Spatial and Financial Neighborhoods.

Dependent Variable: Change in Fertilizer Use Per Plant

Regressor	A		B		C	
	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.
Successful Experiment in Information Neighborhood			0.968	0.169	0.969	0.169
Successful Experiment in Geographical Neighborhood	0.758	0.257	0.063	0.149	0.064	0.148
Successful Experiment in Financial Neighborhood					-0.473	0.222
Constant	1.680	1.411	0.555	1.345	0.720	1.408
Village 1	-10.581	3.365	-4.682	2.237	-5.052	2.286
Village 2	-0.209	1.230	0.966	1.309	0.906	1.327
Wealth	0.050	0.311	-0.405	0.448	-0.439	0.455
Clan 1	-1.179	1.310	-0.027	1.291	-0.486	1.341
Clan 2	-0.749	1.332	0.397	1.287	0.619	1.262
Religion 1	-0.843	1.199	0.742	1.553	1.232	1.477
	n=107		n=107		n=107	

Spatial standard errors calculated as described in Footnote 22. Successful innovations in neighborhood variable as defined in equation (7) with neighborhoods defined using responses to:
"Have you ever gone to ___ for advice about your farm?"

Table 5: Innovations in Fertilizer Use, with aggregate and village-level shocks.

Dependent Variable: Change in Fertilizer Use Per Plant

Regressor	A		B		C		D	
	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.
Successful Experiment in Information Nbhd	0.946	0.148	0.881	0.161	0.909	0.152	0.876	0.162
Successful Experiment in Geographic Nbhd	0.081	0.147	0.136	0.142	0.118	0.151	0.145	0.142
Successful Experiment in Financial Nbhd							-0.290	0.287
Constant	2.115	1.500	1.922	0.812	2.056	1.219	-5.062	2.387
Village 1	-5.588	2.413	-6.943	2.416	-6.649	3.184	5.873	2.295
Village 2	0.485	1.201	-1.122	1.060	-0.840	1.899	6.976	2.428
Wealth	-0.252	0.412			-0.204	0.394		
Clan 1	-0.700	1.231			0.320	1.672		
Clan 2	-0.249	1.266			0.322	1.279		
Religion 1	0.129	1.381			-0.126	1.516		
Round 5	-0.923	1.247	2.474	3.116	2.121	3.339	2.480	3.132
Round 6	-1.356	1.194	-3.734	2.196	-3.457	2.510	-3.791	2.205
Round 7	0.715	2.312	-1.224	1.205	-1.561	1.385	-1.238	1.203
Round 8	-3.524	1.231	-0.927	1.298	-0.912	1.317	-0.938	1.302
Round 9	-2.720	1.406	1.106	1.490	0.626	1.876	1.094	1.491
Round 10	0.098	1.641	1.867	2.584	1.863	2.790	1.851	2.587
Round 11	1.811	1.814	3.457	2.557	3.195	3.114	3.459	2.569
Village 3*Round 5			-5.166	3.272	-4.619	3.837	-5.373	3.340
Village 3*Round 6			2.948	2.697	2.841	2.930	3.296	2.836
Village 3*Round 7			3.719	3.993	3.981	4.039	3.806	3.992
Village 3*Round 8			-5.263	2.609	-5.105	2.938	-5.189	2.604
Village 3*Round 9			-4.834	1.862	-4.202	2.381	-4.524	1.871
Village 3*Round 10			-3.813	2.735	-3.732	2.936	-3.727	2.719
Village 3*Round 11			-2.407	3.395	-2.224	4.320	-2.361	3.411
	n=107		n=107		n=107		n=107	

Spatial standard errors calculated as described in Footnote 22. Successful innovations in neighborhood variables as defined in equation (7) with neighborhoods defined using responses to:
"Have you ever gone to ___ for advice about your farm?"

Table 6: Learning Term Estimates with Alternate Definitions of the Information Neighborhood

Dependent Variable: Change in Fertilizer Use Per Plant

Definition of Information Neighborhood	Individual Covariates*		Village-Round Interactions**	
	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.
Do you know __? (Random Sample of 6 Plots)	0.89	0.18	0.90	0.10
Knowledge of the Existence of the Plot (Random Sample of 6 Plots)	0.95	0.38	0.85	0.24
Talk at Least Once/Week (Random Sample of 6 Plots)	1.03	0.32	0.85	0.23
Reports Learning Interaction (Roster of Contacts Listing of All Significant Learning Exchanges)	0.58	0.23	0.32	0.19
Reports Any Interaction (Roster of Contacts Listing of Transactions and Learning Activities Over 10 Months)	0.77	0.21	0.52	0.17
Predicted Links (Predicted Prob. Of Link >.05)	1.13	0.12	1.07	0.13

Spatial standard errors calculated as described in Footnote 22.
*Individual Covariates specification includes as regressors: successful innovations in the geographical neighborhood, successful innovations in the financial neighborhood, village dummies, wealth, and dummies for religion and extended family.
**Village-Round Interactions specification includes as regressors: successful innovations in the geographical neighborhood, village dummies, round dummies,village 3 - round interactions.

Table 7: Innovations in Profits

Dependent Variable: Change in Profits Per Plant

Regressor	Individual Covariates*		Village-round Interactions**	
	Parameter Estimate	Spatial S.E.	Parameter Estimate	Spatial S.E.
Successful experiment in info. neighborhood	14.21	3.29	10.11	3.58

Spatial standard errors calculated as described in Footnote 22.
Successful innovations in neighborhood variable as defined in equation (7) with neighborhoods defined using responses to "Have you ever gone to ___ for advice about your farm?"

*Individual Covariates specification includes as additional regressors: successful innovations in the geographical neighborhood, successful innovations in the financial neighborhood, village dummies, wealth, and dummies for religion and extended family.

**Village-Round Interactions specification includes as additional regressors: successful innovations in the geographical neighborhood, village dummies, round dummies,village 3 - round interactions.

Appendix Table 1: Probit Model of Information Links

Regressor	Parameter Estimate	Standard Error
Either individual holds heriditary office	-0.367	0.087
Members of the same church	-0.035	0.095
Members of the same clan	0.322	0.079
Both first generation in village	-1.170	0.300
Same gender	0.768	0.088
Have plots w/same soil type	0.086	0.100
Distance between plots	-0.180	0.055

Estimate of Equation (5). Huber standard errors adjusted for possible correlation across observations for a given individual, but not for any spatial correlation.

Appendix Table 2: Descriptive Statistics

Variable	Mean in Estimation Sample	Std. Deviation
Wealth	2.13	2.59
Clan 1	0.35	0.48
Clan 2	0.44	0.50
Religion 1	0.49	0.50
sand	0.48	0.50
loam	0.06	0.23
pH	5.66	0.71
% Organic Matter	2.65	0.83
Fertilizer per plant	2.05	5.76
Profit per plant	253.00	402.00
Indices of Successful Experiments with Fertilizer in Information Nhbd:		
Matched Pair Measures:		
Have you gone to ___ for advice?	0.40	6.32
Predicted links	0.79	6.31
Matched Pairs From Random Sample of 6 Plots:		
Do you know __?	1.29	5.93
Knowledge of the existence of the plot	1.13	6.40
Talk at least once/week	0.62	6.68
Roster of Contacts Measures:		
Significant Conversations Re: Farming	1.51	7.79
All recorded interactions	1.25	7.30
Index of Successful Experiments in Geographic Neighborhood:	1.61	5.48
Index of Successful Experiments in Financial Neighborhood:	0.24	1.10

Appendix Table 3. Correlations Between Neighborhood Regressors

	Advice	Know	Exist	Talk	Learning Interaction	Any Interaction	Predicted	Geographic	Financial
Have you ever gone to ___ for advice re: farm?	1								
Do you know ___? (6 Random Plots)	0.581	1							
Knowledge of Plot's Existence (6 Random Plots)	0.5915	0.7693	1						
Talk at Least Once/Week (6 Random Plots)	0.6468	0.6854	0.893	1					
Reports Learning Interaction (Roster)	0.5223	0.5468	0.753	0.8641	1				
Reports Any Interaction (Roster)	0.5622	0.5928	0.8131	0.9257	0.9706	1			
Predicted Links	0.8386	0.8686	0.7216	0.7184	0.5885	0.6369	1		
Geographic	0.5056	0.5622	0.8042	0.8043	0.708	0.7652	0.5409	1	
Financial	0.0249	-0.0069	0.0067	-0.0092	0.051	0.0719	0.0197	0.1044	1