Measuring Social Networks' Effects on Agricultural Technology Adoption*

Annemie Maertens and Christopher B. Barrett

American Journal of Agricultural Economics, forthcoming

* Annemie Maertens is Assistant Professor at the University of Pittsburgh. Christopher B. Barrett is Stephen B. and Janice G. Ashley Professor of Applied Economics and Management at Cornell University. The data were gathered in collaboration with the ICRISAT and funded through NSF Grant No. 0649330, AAEA and Cornell University.

Technological improvements, perhaps especially in agriculture, drive sustainable advances in labor productivity, incomes, food security and general economic growth. But improved technologies are not adopted immediately, randomly or completely throughout a population. Any firm economic understanding of the diffusion of a new technology therefore depends on understanding the dynamic and cross-sectional patterns of technology adoption. While the noneconomic social sciences have long emphasized social networks' role in the diffusion of new technologies (Rogers 1962), economists by and large focused on access to input (including financial) and output markets, self-insurance associated with farm size and household wealth, and learning from extension services and experimentation (Feder et al. 1985). More recently, economists have started to explore how social networks affect farmers' learning about and uptake of new technologies. The introduction of social networks into technology adoption models allows for a range of potential externalities that matter for policy purposes. Given varied information transmission channels and associated social multiplier effects, which group(s) should one target for extension services or input subsidies in order to yield the biggest or most poverty-sensitive impacts? How effectively do government or firm initiatives encourage uptake if they (at least partly) crowd out informal services provision through existing social networks? Who should bear the costs of initial experimentation in order to enable the broader network to learn about a new technology? The emerging literature on the role of social networks in technology adoption directly speaks to these important issues. This paper briefly reviews this literature, focusing on the measurement of social learning networks (Foster and Rosenzweig 2010 offer a broader review).

A social network is defined by individual members (nodes) and the links among them through which information, money, goods or services flow. The importance of a link is not the same as the frequency of exchange over the link – e.g., Granovetter (1973) finds that 'weak'-low frequency links are more important in the job search process than 'strong'-high frequency links – and, more generally, different links may have different value and behavioral influence. A given link may be unidirectional – i.e., flows are one way only, as from trade publications to farmers or from master farmers to novices – or bidirectional and a network may mix links of both types.

Identifying and measuring the effects of social networks – also called 'social interaction effects' – on technology adoption is not a trivial matter (see, e.g., Manski 1993, Brock and Durlauf 2000). The first challenge involves identifying appropriate reference groups – who is in an agent's network? Is the network defined by prospective/hypothetical or retrospective/activated links? Does one include just direct ('first order') links or also indirect ('higher order') links? Once one identifies the network, obtaining an accurate picture of a farmer's social network from the information contained in a limited sample is not straightforward either. Existing methods of doing so all have shortcomings, as we discuss further below. Second, even if social networks are well-measured, inferring causal social interaction effects from correlations in individuals' behavior is difficult. Within an identified reference group, there almost surely exist correlated attributes among individuals. Agents' behaviors and characteristics affect not only the formation and structure of social networks, they may likewise influence the behaviors of other network members, giving rise not just to changed economic outcomes but also to feedback that causes network structure to evolve endogenously (Barrett 2005, Jackson 2008, Stephens 2009, Chantarat and Barrett forthcoming). So whether due to these matching and selection effects, a common external (agroecological or economic) environment or other confounding factors, spurious correlation in behaviors and outcomes often leads analysts to overstate the importance of social interaction effects. A third challenge arises when agents interact and change behavior simultaneously, generating a 'reflection problem', essentially, making it difficult to separate endogenous from exogenous effects (Manski 1993).

Some theoretical considerations

Theoretical models of social learning differ along the following dimensions: (i) What do farmers value and over what time period? (ii) What type of information does the farmer absorb and from whom? (iii) How does the farmer learn, i.e., how does he update beliefs? (iv) How do beliefs translate into actions? and (v) Do agents interact strategically?

Most economic models of technology adoption assume farmers value a discounted stream of future farm profits. Taking into account attitudes towards (inevitable) risk, the expected discounted utility model is a popular tool, although alternatives exist, such as prospect theory and maximin expected utility theory. Somestudies emphasize other considerations besides the time path of profits, such as behavioral conformity, others' well-being (Barrett 2005, Moser and Barrett 2006, or present a trait-based approach that allows incorporation of individual preferences over traits (Useche, Barham and Foltz 2009). Where applicable, such factors can be taken into account. The farmer's objectives influence which links he monitors and what information he follows. Farmers might learn about the prices of inputs and outputs, about output conditional on input use (Besley and Case 1993), past profitability (Cameron 1999), or the optimal use of inputs under a new technology (Conley and Udry 2010, Foster and Rosenzweig 1995). In the latter case, ‘target input’ models assume that for all farmers the new technology is ‘superior’ to older ones conditional on getting the input choice right. This might be an overly strong assumption when dealing with agricultural technologies, as reflected in the significant disadoption rates seen among many new technologies (Moser and Barrett 2006, Suri 2011).

A typical model assumes that farmers learn by observing others’ experimentation. Compared to learning by doing, this increases uncertainty about information quality, in part due to less accurate knowledge of key complementary inputs (e.g., irrigation, soil quality). Thus farmers engage in `incomplete learning’ and weigh each piece of information proportionately to its value (Conley and Udry 2001). Knowledge of the details of a contact’s action (e.g., area under new technology) or the contact’s information sources (e.g., connection to extension agents) might, in addition, allow the farmer to deduce information about the precision of the contacts’ beliefs. Relatedly, farmers may hold heterogeneous beliefs about the technology such that it does not lend itself to specification using a unique functional form about which agents learn a set of parameters (i.e., global learning). Indeed, farmers’ own narratives of their learning processes are more appropriately described as ‘local learning’ (Conley and Udry 2010, Maertens 2010).

The farmer uses available information to update his beliefs. Most models assume Bayesian updating because it is analytically simple and consistent with the expected utility framework, although more general more ad-hoc learning models are available. Adoption decisions turn on farmers' updated beliefs. But unless learning is an end unto itself, forward-looking farmers might engage in strategic delays, concluding that costs of experimentation exceed the potential benefits of active learning, thereby favoring waiting and (imperfectly) observing others' costly experiments (Foster and Rosenzweig 1995).

Measurement challenges

The empirical challenge of measuring social learning effects is twofold: defining and measuring attributes of the network correctly and gathering adequate data to control properly for factors that might otherwise generate spurious correlation. We focus on the former in this section and address the latter in the next section. The simplest method equates social networks with membership in certain groups, such as a village, caste or gender. This practice is common when lacking explicit network data (Foster and Rosenzweig 1995, Munshi 2004). But this almost surely misrepresents the network and is especially problematic when lacking an exogenous source of inter- or intra-group variation.[1] Alternatively, one could take a census of a village and ask all farmers to list their information contacts. This time-intensive approach might be appropriate in a small and closed village context (Van Der Broeck and Dercon 2011). But it is typically infeasible for larger geographic and social units; and a census of an overly small unit risks artificially truncating the measured social network.

If one decides to sample rather than census, several techniques appear in the literature. A common technique is respondent-driven `snowball sampling' (i.e., subjects recruit further respondents from among their acquaintances). Snowball sampling is useful when one is interested in properties of the network itself, but results in a non-representative sample of the households and thus does not lend itself to conventional inference. Another common approach is to use the 'network within sample' method, asking each farmer about his link to every other person in the sample. However, this technique artificially truncates the network and might result in biased estimates of behavior in the presence of structured networks as unobservables influence both the probability of a link and, independently, the behavior of interest (Santos and Barrett 2008, Chandrasekhar and Lewis 2011). Alternatively, one can ask the farmer to list a certain number of people (typically 5 to 10, or in some cases unlimited) from whom he learns (Bandiera and Rasul 2006). If the survey caps the number of links the farmer can declare, truncation bias in estimates of behavior becomes an issue. In addition, while asking about one's network in an open manner saves time, this method might result in only eliciting the farmer's 'strong' network links, as opposed to 'weak' links, and is likely subject to unobserved heterogeneity in respondents' standards for including – or forgetting to include – links.

This brings us to the importance of how the link question is framed. One inevitably gets different answers depending on whether one asks a farmer whether he knows person X, whether he speaks ('regularly') to person X about farming, etc. It is therefore important to link one's social network question(s) with a social learning model. For instance, in order for a farmer to learn directly from a contact's experimentation he must observe – at a minimum – the technology choice of that contact, and the corresponding output and input application rates. Questions such as 'What is farmer's X's cultivar choice,' might then be appropriate. But even if farmers claim to know contacts' behavior or outcomes, farmers are often ill-informed, forget what they once knew, or project their own beliefs and actions on others (Hogset and Barrett 2010). Although what matters to the farmer is his beliefs, whether correct or incorrect, at the time of decision-making and his certitude about those beliefs, one might want to get sense of their accuracy, which requires information on both the reported and actual behavior and outcomes of one's contacts, which implies that the contacts need to be part of the sample.

This brings us to the last sampling technique: `random matching within sample', where each farmer is matched with a certain number (typically 5 to 10) of randomly drawn individuals from the sample and, for each match, one elicits the details of the relationship between the farmer and the match, including possibly the farmer's knowledge about the match's farming activities and outcomes (Conley and Udry 2010, Santos and Barrett 2010, Maertens 2010). Santos and Barrett (2008) show, using Monte Carlo simulation on a network of Ethiopian herders, that this method outperforms the `network within sample' method. While this method has the advantage that it can be integrated in a time efficient manner within an existing sample, and one can use predicted links as generated regressors in subsequent regression analysis, one needs to be wary in the presence of certain network structures. If the sample omits a key network node, i.e., someone with many links compared to others, the resulting omitted variable bias can be substantial.

Social networks in rural India

We briefly illustrate some of the preceding points using data collected on social networks among 246 International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) Village Level Studies (VLS) farming households in India's semi-arid tropics in 2007-08, focusing on the structure of the network, framing of questions and the nature of learning. The average household in the sample is poor, with just 6 acres of land for 4.5 household members and little education. About half of the sample households cultivate cotton, and among these 64% cultivate *Bacillus thurigiensis* (Bt) cotton, a new variety that requires fewer pesticides and increases expected yields. Bt cotton has been the main object of learning among the surveyed farmers since 2002. We collected social network data using the random matching within sample technique, in which each respondent was randomly matched with six VLS respondents and a set of four fixed `progressive farmers'. These 'progressive farmers' represent the most successful farmers in each village; we expected them to play a central role in the diffusion of information (Rogers 1962).

Denote the matched individual by X. One advantage of this method is that the respondent is less likely to 'forget' about a contact. Indeed, we find that, when comparing the answers to the open question: 'who would you go to for advice in case of problems with your cotton crop?' with the answers to the closed question: 'would you go to X for advice in case of problems with your cotton crop?', on average, only 25% of the matches whom the respondent would approach in the closed questions are mentioned when being asked exactly the same question in an open-ended manner beforehand.

On average, 93% of the matches were known to the respondent. Conditional on knowing cotton farmer X, the respondent thinks he knows the cultivar choice (of farmer X) for, on average, 75% of the matches, the yield for, on average, 41%, the pesticide use for, on average, 40%, soil characteristics for, on average, 92% and irrigation status for, on average, 75% of the matches. Yet the best guess of the farmer in this regard is correct only 82% (cultivar), 17% (yield), 32% (pesticide), 38% (soil), and 51% (irrigation) of the time, respectively. These statistics suggest that the information farmers have about each other's production activities is often incomplete, inaccurate or both. If the farmers are aware their imperfect information, optimal learning from other farmers' experimentation implies that such information is weighted to take into account this uncertainty. Alternatively, it might be possible that the respondents are over-confident when being questioned about their links with each other as they are reluctant to admit social isolation. But that too implies a need to weight responses in modeling learning.

In addition, farmers are not always aware of each other's networks, i.e., with whom their learning contacts themselves link, which limits what the farmer can learn about his contacts' beliefs. In 20% of the matches (for which we have information on both sides), the respondent incorrectly assumes that the knowledge relationship (with regard to yield and pesticide use) is symmetric. The lack of symmetry in networks – equivalently, the importance of unidirectional links – is underscored by the fact that for 45% of the matches with progressive farmers, the progressive farmer says he never speaks with the respondent, while the respondent has a very different perception of the relationship.

Table 1 presents Probit regression results relating the existence of a 'learning link' – defined as the respondent thinks he knows the match's cultivar choice, yield outcome and pesticide use – to characteristics of the respondent, the match, and their relationship. We present the results separately for the three villages, as the underlying social network structures differ sharply (a Chow test rejects the null of equality of the coefficients across the three villages at the 1% level). Belonging to the same sub-caste is associated with an increased likelihood of a learning link. Location matters as well. Living near another farmer, having a field near another farmer's field, or merely passing by another farmer's field on a regular basis all increase the likelihood of a link, by 20 to 50%.

Learning network structure nonetheless clearly differs among even seemingly very similar villages. In Kanzara, social networks are structured along income and land classes, but cut across educational levels while in Aurepalle, social networks are structured along educational levels and land classes, but appear to cut across income levels. The apparent heterogeneity of what causes farmers to link with each other to learn about farming serves as a caution about imposing too much *a priori* structure on the link formation and about assumptions of global learning.

This simple regression likewise demonstrates the prospective hazards of correlated behaviors due to unobserved environmental conditions or farmer attributes, such as soils or risk preferences. In Aurepalle, farmers who have similar soil conditions are, on average, about 15% more likely to be in each other's learning network, consistent with a local learning model in which the production of cotton depends on complementary inputs, including soil quality. Similar conclusions can be drawn for Kanzara based on the average value of the land (in Rs/acre) as a measure of soil quality. Meanwhile, in Kanzara and Aurepalle, farmers are more likely to link with a match who is less risk averse, consistent with a model of learning from less risk averse experimenters, while in Kinkhed the opposite is true. The prospect that any of these relationships are causal – shared soils or different risk preferences drive link formation – serves as a caution about spuriously correlated behavior within networks in adoption regressions.

Another source of correlated behavior is the unidirectional links between the progressive farmers and the other farmers in the village. In the three villages, farmers were significantly more likely to know a farmer, and to be aware of the details of his farming activities, if the match is a progressive farmer. In addition, farmers might receive information through visits of company agents to the village, government extension agents and conversations with input dealers in the nearest urban hub (see also Matuschke and Qaim 2009). The respondents had heard from, on average, 0.9 outside sources in the last seven years about Bt cotton, and found this information 'useful' to 'very useful' in 75% of the cases, on average. These findings underscore the importance of including all sources of information in any study of social learning.

Discussion

Economists increasingly appreciate the important role social networks play in mediating the diffusion of agricultural innovations. But this literature remains distinctly underdeveloped. We see several important areas in need of attention in future research.

First, closer attention needs to be paid to the manner in which social network structure data are collected. Second, more detailed data collection is necessary to control for potentially confounding variables that may lead to spurious correlation among agents' behaviors – georeferencing using GPS to link to biophysical information available in soils, climate and agronomic databases; behavioral experiments to elicit otherwise-unobservable parameters for risk and time preferences, trust, etc.; elicitation of agents' subjective beliefs about different technologies and their traits, market prices, etc. Panel data collection on social networks, information flows and these other typicallyunobservable variables can then be used to shed light on endogenous network formation, in particular, how agents' network structures respond to the flow of information. Quasi-panel data might be used as an imperfect substitute in some cases, but note that some variables are intrinsically hard to recall, such as growing conditions, income, labor availability, beliefs, and perhaps even networks themselves. Third, using these data, researchers could test a variety of learning and social interaction models against one another, instead of imposing a particular model on the data. This can shed light on the sequencing of interactions and the nature of information flows. Explicitly modeling of social learning as well as of the strategic interactions among farmers reduces the reflection problem and related endogeneity problems surrounding estimated interactions effects. Fourth and finally, but perhaps most importantly, economists need to relax the widespread assumption that social interactions' effects on technology adoption reflect 'learning' as distinct from non-learning social interaction effects due to, for example, coordination in marketing or water management, mutual assistance in labor or insurance, informal credit, or the non-material values of conformity or imitation. Disentangling behaviors that are nested within complex social structures is a daunting challenge but an important one to tackle in the next generation of economic research on the effects of social networks on agricultural technology adoption and other high-impact activities.

References

Bandiera, O. and I. Rasul. 2006. "Social Networks and Technology Adoption in Northern Mozambique." *Economic Journal* 116:869–902.

Barrett, C.B., ed. 2005. *The Social Economics of Poverty: on Identities, Groups, Communities and Networks* (London: Routledge).

Barrett, C.B. and M.R. Carter. 2010. “The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections.” *Applied Economics Perspectives & Policy* 32:515-48.

Besley, T. and A. Case. 1993. “Modeling Technology Adoption in Developing Countries.” *American Economic Review* 83:396-402.

Brock, W. and S. Durlauf. 2000. “Interactions-Based Models,” in J. Heckman and E. Leamer, eds., *Handbook of Econometrics, vol. 5* (Amsterdam: North-Holland).

Cameron, L.A. 1999. “The Importance of Learning in the Adoption of High-Yielding Variety Seeds.” *American Journal of Agricultural Economics* 81:83-94.

Chandrasekhar, A.G. and R. Lewis. 2011. "Econometrics of Sampled Networks." Mimeo.

Chantarat, S. and C.B. Barrett. Forthcoming. “Social Network Capital, Economic Mobility and Poverty Traps,” *Journal of Economic Inequality*.

Conley, T.G. and C.R. Udry. 2001. “Social Learning through Networks: the Adoption of New Agricultural Technologies in Ghana.” *American Journal of Agricultural Economics* 83:668-73.

Conley, T.G. and C.R. Udry 2010. "Learning about a new technology: pineapple in Ghana." *American Economic Review* 100:35-69.

Feder, G., R.E. Just, and D. Zilberman. 1985. “Adoption of Agricultural Innovations in Developing Countries: A survey.” *Economic Development and Cultural Change* 33:255-98.

Foster, A.D. and M.R. Rosenzweig. 1995. “Learning by Doing and Learning from Others: Human Capital and Technical Changes in Agriculture.” *Journal of Political Economy* 103:1176-1209.

Foster, A.D. and M.R. Rosenzweig. 2010. “Microeconomics of Technology Adoption.” *Annual Review of Economics* 2:395-424.

Granovetter, M.S. 1973.”The Strength of Weak Ties.” *American Journal of Sociology* 78: 1360-80.

Hogset, H. and C.B. Barrett. 2010. “Social Learning, Social Influence and Projection Bias: A Caution on Inferences Based on Proxy-Reporting of Peer Behavior.” *Economic Development and Cultural Change* 58: 563-89.

Jackson, M.O. 2008. *Social and Economic Networks* (Princeton, NJ: Princeton University Press).

Maertens, A. 2010. *Social Networks, Identity and Economic Behaviour: Empirical Evidence from India*. Cornell University Ph.D. dissertation.

Manski, C. 1993. “Identification of Endogenous Social Effects: The Reflection Problem” *Review of Economic Studies* 60:531-42.

Matuschke, I. and M. Qaim. 2009. “The Impact of Social Networks on Hybrid Seed Adoption in India.” *Agricultural Economics* 40:493-505.

Moser, C.M. and C.B. Barrett. 2006. “The Complex Dynamics of Smallholder Technology Adoption: The Case of SRI in Madagascar.” *Agricultural Economics* 35:373-88.

Munshi, K. 2004. “Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution.” *Journal of Development Economics* 73:185-213.

Rogers, E. 1962. *Diffusion of Innovations* (New York, Free Press).

Santos, P. and C.B. Barrett. 2008. “What Do We Learn From Social Networks When We Only Sample Individuals? Not Much.” Mimeo.

Santos, P. and C.B. Barrett. 2010. “Identity, Interest and Information Search in a Dynamic Rural Economy.” *World Development* 38:1788-96.

Stephens, E. 2009. “Feedback Relationships between New Technology Use and Information Networks: Evidence from Ghana.” Pitzer College unpublished manuscript.

Suri, T. 2011. “Selection and Comparative Advantage in Technology Adoption”, *Econometrica*, 79: 159–209.

Van den Broeck, K., and S. Dercon. 2011. "Information Flows and Social Externalities in a Tanzanian Banana Growing Village." *Journal of Development Studies* 47:231-52.

Useche, P., B. Barham and J. Foltz. 2009. "Integrating Technology Traits and Producer Heterogeneity: A Mixed-Multinomial Model of Genetically Modified Corn Adoption." *American Journal of Agricultural Economics* 91:444–61.

Table 1: Correlates of the learning network in three Indian villages

Probit regression with dependent variable: presence of a "learning link" between respondent and match

Village:	Aurepalle		Kanzara		Kinkhed		All villages	
	dF/dX	Error	dF/dX	Error	dF/dX	Error	dF/dX	Error
Relative risk preferences	0.107***	(0.039)	0.111***	(0.034)	-0.092**	(0.039)	0.029	(0.021)
Similar soil conditions	0.146***	(0.056)	0.005	(0.063)	0.027	(0.063)	0.064*	(0.038)
Belong to same organization	0.102	(0.068)	0.081	(0.065)	0.107*	(0.064)	0.125***	(0.039)
Live in same neighborhood	0.178**	(0.087)	0.075	(0.113)	0.167**	(0.060)	0.150***	(0.052)
Pass by X's field when going to field	0.130*	(0.074)	-0.013	(0.100)	-0.272**	(0.146)	0.028	(0.055)
X's field close to respondent's field	0.199**	(0.099)	0.334**	(0.141)	0.163**	(0.070)	0.184***	(0.056)
Belong to same sub-caste (*jati*)	0.216***	(0.083)	0.178**	(0.084)	0.161*	(0.070)	0.186***	(0.050)
Have same family name	-0.016	(0.141)	0.155	(0.121)	0.036	(0.210)	0.125	(0.084)
Education of HH head (sum)	0.026***	(0.007)	-0.004	(0.011)	0.003	(0.008)	0.006	(0.004)
Education of HH head (diff)	-0.015**	(0.007)	0.027**	(0.011)	0.005	(0.008)	-0.004	(0.004)
Income (10,000 Rs) (sum)	0.017***	(0.005)	-0.006**	(0.003)	0000	(0.005)	-0.004**	(0.002)

Income (diff)	0.015***	(0.005)	-0.005	(0.003)	0.002	(0.005)	0.001	(0.002)
Land (acres) (sum)	0.000	(0.006)	0.008**	(0.005)	-0.004	(0.003)	0.002	(0.002)
Land (acres) (diff)	-0.014**	(0.007)	-0.016***	(0.005)	0.000	(0.004)	-0.002	(0.002)
Land value (10,000 Rs/acres) (sum)	0.002	(0.002)	0.001	(0.005)	0.018**	(0.008)	0.000	(0.001)
Land value (diff)	-0.005*	(0.003)	-0.029**	(0.013)	-0.001	(0.020)	-0.002	(0.003)

Notes: *** p<0.01; ** p<0.05; * p<0.1; standard errors (clustered at the individual's and the match's level) are reported in parentheses next to the average marginal effect estimates. Controls for sum and difference of number of household members, number of adults, value of machinery, age of household head, and irrigation status. "sum" and "diff", respectively, refer to the sum and difference of the attributes of respondent and match.. Relative risk preferences measured as the answer to: How risk averse are you compared to X (X takes much more risk(1) to I take much more risk (5)). Similar soil conditions is a dummy variable which takes the value of 1 if respondent believes he and X have a similar type of soil. Income refers to the rainy season (*kharif*) income. Total number of observations = 1096.

[1] One could imagine selecting a random subset of groups or group members and providing this set with information about a new technology or a financial incentive to induce adoption or randomizing rollout of a new technology so as to introduce exogenous variation experimentally. However, essential heterogeneity might affect how farmers both interpret and react to the treatment, thereby undermining both the external and internal validity of the experimental design (Barrett and Carter 2010).