Technology Diffusion, Outcome Variability, and Social Learning: Evidence from a Field Experiment in Kenya

Andrew Crane-Droesch

This article explores the mechanisms through which social learning mediates technology diffusion. We exploit an experiment on the dissemination of biochar, a soil amendment that can improve fertility on weathered and/or degraded soils. We find that social networks transmit information about the average benefits of adoption, but also its risk, and that observed variability inhibits uptake to a greater degree than positive average results engender it. Paradoxically, this relationship is stronger among networks that do not discuss farming, but disappears among farmer networks that do. This is resolved with a simple model of social learning about conditional, rather than unconditional benefit distributions. As farmers observe factors associated with outcomes in their networks, they constrain the distribution of their own potential outcomes. This conditional distribution diverges from the unconditional distribution that the econometrician observes. We conclude that social learning is characterized by implicit model-building by sophisticated decision makers, rather than simple herding towards observed good results.

Key words: Agriculture, development, Kenya, learning, risk, semiparametric econometrics, social networks, technology adoption.

JEL codes: O33, O13, Q12

Dissemination of novel agricultural technologies in sub-Saharan Africa remains a challenge. Productivity has grown very little in the region since independence, while the population remains largely rural and poor (World Bank 2014). Slow diffusion of higher-productivity agricultural technologies explains much of this stagnation, which is particularly stark in contrast to the massive productivity gains experienced by most of the rest of the world during the same period. While heterogeneous profitability accounts for part of this slow rate of diffusion (Marenya and Barrett 2009a; Suri 2011), the relationship between uptake and profitability is not always straightforward (Duflo, Kremer, and Robinson 2011b).

Social learning plays an important role in technology diffusion. Despite challenges in distinguishing it from imitation and correlation (Manski 1993), its presence has been shown both to affect the rate of technology dissemination (Foster and Rosenzweig 1995), and change management practices for already-adopted technologies (Conley and Udry 2010). The efficacy of social learning has been shown to be mediated by environmental heterogeneity (Munshi 2004), and its effect has been shown to be negative where private benefits are negative (Kremer and Miguel 2007). The literature also shows that social learning is not always effective at speeding the rate of technology diffusion, even when technologies appear profitable (Duflo, Kremer, and Robinson 2008). Finally, social learning effects can be nonlinear (Bandiera and Rasul 2006), and mediated by the characteristics of social linkages

Andrew Crane-Droesch is a research economist for the Economic Research Service, USDA. This project received support from USAID Development Innovation Ventures (grant number AID-OAA-F-00016), the U.S. Government's Feed the Future program through a Borlaug Fellowship in Global Food Security, the International Food Policy Research Institute, the Center for Effective Global Action, and Innovations for Poverty Action. The views expressed are those of the author and should not be attributed to the Economic Research Service or the USDA. Correspondence to be sent to: andrew.crane-droesch@ers.usda.gov.

Amer. J. Agr. Econ. 0(0): 1–20; doi: 10.1093/ajae/aax090

(Banerjee, et al. 2013; Cai, De Janvry, and Sadoulet 2015).

This article builds on the above literature in two ways. The first contribution is to show that information about risk is transmitted socially. It is well-known that poor farmers in sub-Saharan Africa (SSA) tend to be risk-averse, and that this mediates their uptake of new technologies (Dercon and Christiaensen 2011; Karlan, et al. 2014). Using exogenous shocks to information sets, we show that adoption is sensitive to the mean and variance of in-network results, but more sensitive to variance than to the mean.

The second contribution is to characterize how social learning operates. We begin with a simple model of learning about outcome distributions, in which farmers observe outcomes together with factors that mediate them. The mediators are then used to form a model for the outcome, which provides an updated conditional distribution. The farmer then bases their adoption decision on this unobservable conditional distribution.

The model is supported by the data. Among farmers who share more information with one another, adoption is sensitive to the mean and variance of results among linked farmers. Among farmers who share less, adoption is insensitive to these moments.

Identification of social learning about the means and variances of observed outcomes is challenging. An individual's characteristics will determine who is in their network, and different sorts of farmers are likely to have different outcomes to the adoption of any given technology. The same characteristics that determine social network composition probably also mediate adoption propensity. Observed outcomes within a social network are thereby endogenous to the decision to adopt.

However, the present context offers a strategy for identifying the causal effect of the mean and variance of the observed benefit. The technology in question —biochar—tends to cause larger yield increases where baseline yields are low: it usually works better on poor soils than on highly fertile ones (Crane-Droesch, et al. 2013). If low-yielding farmers are inherently less likely to adopt, and if they tend to know other low-yielding farmers—who will tend to get stronger response to biochar adoption—then a naive estimate of the effect of the observed average result within a social network will be biased downward. Likewise, if more gregarious farmers are inherently more likely to adopt (perhaps they are wealthier and/or more innovative, etc.), and if they tend to see a wider range of outcomes than the less gregarious, then a naive estimate of the effect of the standard deviation of observed outcomes on uptake will be biased upward. This article uses data from an experiment in which inducements to adopt biochar were randomly assigned. Together with a fair degree of predictability in outcomes to adoption, these inducements are used to construct instrumental variables for the outcome distribution that each farmer observes within their social network.

The experiment underlying this article focused on the dissemination of biochar (Lehmann and Joseph 2009). Biochar is a charcoal-like soil amendment that can increase crop yields on poor and/or degraded soils in the first year after soil application and beyond (Crane-Droesch, et al. 2013). Given that weathered soils are common in the global south (Sanchez 1977), and that soil fertility is among the constraints to agricultural development (Sanchez 2002; Vanlauwe and Giller 2006), biochar is potentially useful in agriculture in the developing world. Unlike fertilizer, biochar is not directly taken up by the plant. Rather, it mediates soil quality by retaining nutrients and water, providing habitat for soil biota, and improving soil structure and pH. Due to its chemical structure, (Keiluweit, et al. 2010) the carbon in biochar decomposes to carbon dioxide much more slowly than the plant matter from which it was formed (Hammes, et al. 2008; Singh, et al. 2012; Zimmerman 2010; Kuzyakov, Bogomolova, and Glaser 2014). As such, biochar has been suggested as a means of mitigating climate change by sequestering stabilized carbon in soils (Lehmann 2007), potentially at net-negative cost if crop yield increases are sufficiently large.

This article provides two modest methodological contributions. First, it extends generalized additive models (GAMs) (Hastie and Tibshirani 1986)—a semiparametric regression framework—into the context of discrete choice methods for models with endogenous variables. Second, it provides a refinement to methods for computing marginal effects within the Special Regressors framework of Lewbel (2000) and Dong and Lewbel (2015).

Conceptual Framework

This section develops a model of learning about the outcomes of a technology with varying benefits, and builds off of the basic two-moment representation of the utility function (Saha 1997) with respect to the potential outcomes of adoption of a new technology:

$$(1)\quad U(adoption) = \mu^{\theta_1} - \sigma^{\theta_2}$$

where μ and σ are beliefs about the expectation and standard deviation of the profitability (ΔP) of adoption, while the parameters θ_1 and θ_2 define risk preferences.

If farmers are risk-averse and form beliefs about the mean and variance of ΔP that correspond to their in-network sample moments, then our first testable implication is, quite simply: *adoption will be more likely when farmers in their network experience large average yield increases, but lower where yield changes in their network are variable.*

Learning about Mediators of Outcomes

However, farmers observe more than just outcomes; they also observe mediators of those outcomes, if imperfectly. This section describes the process by which those mediators might enter into the learning process.

Begin with a simple example to fix ideas. A farmer knows five others who used biochar in the previous season. Two of them own cows and apply manure to their fields, while three of them do not. None applied biochar to their whole plot, but rather applied biochar only to a portion. The two with the cows see small differences between the biochar and non-biochar sub-plots, while the other three saw large differences. If our farmer does not know about their fellows' manuring practices, then biochar's benefit appears to be mediocre and variable. If, however, our farmer is aware, the benefit becomes a *conditional* function of whether or not the fields are manured. And given this mediator, the conditional variance—and thereby the perceived riskiness—is much smaller.

More formally, consider yield as a complex function of many factors, $\boldsymbol{X}$. These may include practices, inputs, soil characteristics, weather, etc.—anything that influences crop yields. We make no assumptions about how the new technology enters the production function, but we model it as binary $B \in \{0, 1\}$. We therefore have

$$(2)\quad Y = \mathcal{G}(\boldsymbol{X}, B).$$

This is equivalent to

$$(3)\quad Y = \begin{cases} \mathcal{G}^0(\boldsymbol{X}) & \text{if } B{=}0 \\ \mathcal{G}^B(\boldsymbol{X}) & \text{if } B{=}1. \end{cases}$$

Change in yield in response to adoption is thus

$$(4)\quad \begin{aligned} \Delta Y &= (\mathcal{G}^B(\boldsymbol{X}) - \mathcal{G}^0(\boldsymbol{X})) \\ &\equiv \mathcal{F}(\boldsymbol{X}) \end{aligned}$$

while change in profit is

$$(5)\quad \Delta P = p_Y \mathcal{F}(\boldsymbol{X}) - p_B - p_\Delta$$

where p_B is the direct cost of adoption and p_Δ are ancillary costs of adoption such as required changes in labor and other complementary inputs.

The production function is deterministic; if the farmer knows $\mathcal{F}$ and observes $\boldsymbol{X}$ perfectly, then potential change in yield is known with certainty. The farmer i can predict their own yield change by applying $\mathcal{F}$ to their own circumstances $\boldsymbol{X}_i$, computing $\mu_i = p_Y \mathcal{F}(\boldsymbol{X}_i) - p_B - p_\Delta$, while $\sigma = 0$. Adoption occurs if $\mu_i > 0$, all else being equal.

This is clearly unrealistic. Individual farmers do not begin knowing $\mathcal{F}$, while their knowledge of $\boldsymbol{X}$ is incomplete at best. While they observe yields, they do not know which $\boldsymbol{X}$'s are relevant, nor do they perfectly observe them among their social contacts. Furthermore, time-varying elements of $\boldsymbol{X}_i$ (such as weather) may have been observed in the past, but canot be known for the future. As such, the production function appears stochastic.

Consider the extreme case where yields are observed, but no $\boldsymbol{X}$'s are observed. Given uninformative priors, the best that the farmer can do is to operate under the supposition that their expected yield change corresponds to the average yield change that they observe. Denoting the contacts of i as $\boldsymbol{A}_i$ – the ith row of the social network adjacency matrix $\boldsymbol{A}$ – the *unconditional* moments of the profit distribution are thus $\mu_i^U = p_Y \bar{\Delta} Y_{\boldsymbol{A}_i} - p_B - p_\Delta$ and $\sigma_{\boldsymbol{A}_i}^U = p_Y \sqrt{(\Delta Y_{\boldsymbol{A}_i} - \bar{\Delta} Y_{\boldsymbol{A}_i})^2}$. These enter into the utility function (1), and the farmer makes the adoption decision consistent with their risk preferences.

In between these two extremes, consider a farmer who observes $\boldsymbol{X}_{\mathbf{A}_i}$ imperfectly. Instead, they observe $\mathcal{X}_{\mathbf{A}_i} = h(\boldsymbol{X}_{\mathbf{A}_i})$, where $\forall x \in \boldsymbol{X}$,

$$(6)\quad h(x) = \begin{cases} x & \text{with probability } p_x \\ \text{NA} & \text{with probability } 1 - p_x \end{cases}$$

where NA ("not available") denotes a missing value. In other words, farmers observe only a subset of variables, and not all cases of each observed variable. Using data that they do observe in their network, $[\Delta Y_{\mathbf{A}_i}, \mathcal{X}_{\mathbf{A}_i}]$, the farmer then implicitly fits the model

$$(7)\quad \Delta Y_{\mathbf{A}_i} = \mathcal{F}(\mathcal{X}_{\mathbf{A}_i}) + u\,1.$$

Estimates of yield change given a farmer's circumstances are thus $\mathbb{E}[\Delta Y_i|\boldsymbol{X}_i] = \hat{\mathcal{F}}(\boldsymbol{X}_i)$. The estimated equation (7) allow the farmer to (implicitly) compute *conditional* moments of the profitability distribution $\mathbb{E}_i[\mu_i|\boldsymbol{X}_i] \equiv \hat{\mu}_i = p_Y\hat{\mathcal{F}}(\boldsymbol{X}_i) - p_B - p_\Delta$, and $\hat{\sigma}^2 = var(\hat{u})p_Y$.

Assuming that the farmer's implicit estimator of $\mathcal{F}$ explains any of the variance in $\Delta Y_{\mathbf{A}_i}$, the unconditional standard deviation will always be greater than the conditional for a given set of observations, while the conditional vs unconditional estimates of the mean will generally be different.

However, the econometrician observes neither the farmer's information set nor their reasoning process. Instead, only the network and the resultant unconditional moments are observed. These moments are not generally those that the farmer uses to make their decision.

This suggests a testable proposition: *The relationship between adoption and unconditional outcome distribution will be stronger among farmers who do not share as much information.* This is because the unconditional moments of the benefit distribution (observed by the econometrician) will be closer to the implicit parameters upon which the farmer bases their decision.

Incorporating covariates, adoption can thus be represented as

$$(8)\quad \text{Adopt}_i = \boldsymbol{1}(\alpha + \hat{\mu}_i^{\theta_1} - \hat{\sigma}^{\theta_2} + \boldsymbol{Z}\Gamma + \epsilon > 0)$$

where $\boldsymbol{Z}\Gamma$ and $[\alpha, \epsilon]$ represent the effects of other factors—observable and unobservable, respectively—on the adoption decision.

If, for convenience, we assume that the farmer's risk preference can be represented by constant average risk aversion, we have that $\theta_1 = 1$ (following Saha 1997), which allows for additive separability in the terms comprising $\hat{\mu}_i$. This allows us to write the following linear-in-parameters approximation to equation (8)

$$(9)\quad \text{Adopt}_i = \boldsymbol{1}(\alpha + \beta_1\hat{\Delta}Y_i + \beta_2\hat{\sigma} + \boldsymbol{Z}\Gamma + \epsilon > 0)$$

where a farmer's expected yield change $\hat{\mathcal{F}}(\boldsymbol{X}_i) \equiv \hat{\Delta}Y_i$, and incorporating p_B and p_Δ into $\boldsymbol{Z}$. If the farmer is risk-averse, then β_1 will be positive and β_2 will be negative.

Denote the differences between the conditional and unconditional mean and variance d_μ and d_σ. Replacing the conditional moments in equation (9) with their unconditional counterparts yields

$$(10)\quad \text{Adopt}_i = \boldsymbol{1}(\alpha + \beta_1\bar{\Delta}Y_{\mathbf{A}_i} + \beta_2\sigma^U_{\mathbf{A}_i} + \boldsymbol{Z}\Gamma + \epsilon + \beta_3 d_\mu + \beta_4 d_\sigma > 0).$$

The correlation between $\bar{\Delta}Y_i$ and d_μ is ambiguous, depending on a farmer's characteristics $\boldsymbol{X}_i$ and their estimator $\mathcal{F}$. At minimum then, estimates of equation (10) that omit d_μ will suffer from attenuation in estimates of β_1—driving it toward zero from above. The correlation between σ^U and d_σ will generally be positive, while β_2 will presumably be negative. Omission of d_σ will therefore bias estimates of β_2 upward, making it less negative.

However, if the network of reference is one where p is very low—farmers do not observe the mediators of their neighbor's outcomes—then each of the omitted variables will approach zero and equation (10) may be consistently estimated by any strategy that accounts for the endogeneity in observed outcomes.

[1] Assume that the farmer implicitly estimates $\mathcal{F}$ in a manner analogous to expectation-maximization (Dempster, Laird, and Rubin 1977) or multiple imputation (Rubin and Schenker 1986). In other words, farmers are able to form conditional expectations in the presence of incomplete data by forming conditional probability distributions over missing values of variables that they do not completely observe.

The linear structure (in the index) of equation (10) is, for convenience, in exposition of the conceptual model only. Our empirics are semiparametric, using additive models that relax linearity while retaining additivity. As such, our predictions on the signs of coefficients are equivalent to predictions about whether a function is increasing or decreasing. Furthermore, the semiparametric approach allows us to recover the nonlinearity in σ^{θ_2}. Similarly, relaxing our CARA assumption implies that we approximate $(p_Y\bar{\Delta}Y - p_B - p_\Delta)^{\theta_1}$ with $f_1(p_Y\bar{\Delta}Y) + f_2(p_B) + f_3(p_\Delta)$. In fact, this additive approximation may be superior inasmuch as the value of expected yield increases may be separable from the costs of adoption if farmers are cash/credit constrained.

What about the Number of Adopters?

The number of adopters in the social network plays a key role in the classic target-input model of learning about a new technology (e.g., Foster and Rosenzweig 1995), by constraining a farmer's estimate of the variance of the expected benefit. We argue that it may also play a role in shaping the conditional variance of the absolute benefit, rather than just the variance of the expectation of the benefit.

If farmers with more adopters in their network observe a larger support in $\boldsymbol{X}$, then their estimates of the conditional mean $\hat{\mathcal{F}}$ will tend to improve, explaining more of the variance in ΔY. The farmers' estimate of the conditional variance will thus shrink, and they will be more likely to adopt, all else being equal. This would generally happen in situations where farmers are sharing information on the mediators of the results that they observe.

This dynamic does not exclude frameworks such as those of Munshi (2004), Bandiera and Rasul (2006), Conley and Udry (2010), or Foster and Rosenzweig (1995), who explain the effects of the number of adopters variously by free-riding on externalities to experimentation, and learning about the variance of an expected benefit. Neither does it exclude non-learning channels such as social pressure.

Background and Context

The predictions from the conceptual model are applied to data from a field experiment on biochar dissemination. The experiment underlying this article originally sought to test biochar's profitability under smallholder management, and to compare the efficacy and complementarity of various inducements aimed at stimulating uptake. The experiment began by randomly allocating demonstration plots at the farmer-level. This was followed by two rounds in which all farmers in the study were given the opportunity to purchase biochar. This purchasing component had additional layers of randomization embedded in it, randomizing both the price and also an offer to take biochar freely and pay later conditional on good results. As will be shown below, the main experimental results were disappointing. While biochar appeared profitable, uptake in the absence of steep subsidies was quite low, and no clear social influence effect was observed. This article is motivated in part by a desire to explain this apparent puzzle—low adoption despite high average profitability and/or heavy subsidies.

The Study Region

The experiment was conducted in rural areas surrounding Bungoma, western Kenya. The region receives about 1,600mm rainfall annually, distributed bimodally in two rainy seasons. Known locally as the "long rains" and the "short rains," they occur approximately March to June and August to November. Planting and harvesting dates vary farmer to farmer.

Soils in the region are relatively weathered (Sombroek et al. 1982), and agriculture is dominated by maize and sugarcane production. Most maize varieties are hybrids, though there is substantial heterogeneity among the hybrid maize cultivars on the market and in use in the region (Tjernström 2015). Maize is grown mostly for consumption, and most households are net consumers. Sugarcane is a prominent cash crop. Most farmers plant two maize crops per year, and yields in the short rains are typically half of those in the long rains. Population density is higher than much of the rest of rural Africa, averaging 100–250 people/km^2. Average farm sizes are less than 1 hectare, and fallowing—once common in the area—is now rarely practiced.

Sample Selection. Project zones were selected by taking a number of random draws from a circle centered on Bungoma. After removing areas with some prior exposure to biochar, the first four of those remaining were selected.[2] A $2km^2$ box was drawn around these points, and households within the zone defined by the box were identified using satellite images and field visits. Figure A.1 gives the relative locations of the four zones, and the households within them.

Summary statistics are provided in table 1. The sample is composed primarily of smallholder farmers with low yields and little other income. Women are the main agricultural decisionmakers in about two-thirds of households. Fertilizer use is generally low—the average farmer spends just over \$78/ha/season, while government-recommended application rates for maize tend to cost around \$200/ha/season. About one-quarter of respondents were members of One Acre Fund, a non-governmental organization (NGO) that provides agricultural inputs on credit, while one-third reported that they had ever produced charcoal—an income-generating activity commonly associated with poverty.

Small farms and a dense population may facilitate social learning, both by increasing the number of links to and from each farmer, and by making it possible for farmers to observe the yields of people who they know, but with whom they do not neccesarially speak with about farming.

Project Description

A timeline of the project's activities is given in figure A.2, and described below. Project activities followed the agricultural calendar, which largely follows the two rainy seasons. Because the seasons are close together, many agricultural and project activities overlapped.

A baseline survey collected demographic data, as well as recall-based crop yield data for 3 years (6 seasons) beforehand. Because we are interested in agricultural decision making, we define the "household" unit as groups of people who farm separately.[3] In total, 1,115 households were included in our initial sample, which was intended as a census of the four zones.

Demonstration plots. Beginning in August 2013, 75 farmers were selected at random to receive biochar demonstration plots. Participating farmers were told that biochar might improve their crop yields, but that the demonstration plot would allow them (and their neighbors) to judge for themselves. Project staff asked participating farmers to identify a prominent one-eighth acre (.05 ha) portion of their farm—in the sense that it was relatively more visible to passers-by than other portions—to comprise the plot.

Plots were then divided; half was to get biochar, while the other half would not. Which half got biochar was decided with a coin flip. On the biochar half, project staff applied 25kg of biochar, which is equivalent to an application rate of 0.5 tons per hectare. The biochar was produced using a top-lit updraft gasifier (Brown 2009), from dried maize stalks and sugarcane leaves. In order to facilitate passers-by learning about the benefits of biochar, plots were demarcated using twine and marked with signs indicating which had biochar and which did not. Farmers were asked to treat both halves identically with respect to inputs and labor. No farmers reported unequal input use. At harvest in December, project staff measured yields of dried maize on all demonstration plots using a scale (see figure A.3). On average, biochar plots performed substantially better than non-biochar plots, but yield increases were only sufficient to make biochar profitable under mild subsidies.[4,5]

First stales offer. Beginning in late January 2014, enumerators offered random sales offers to all households. About 20% of the farmers from the original census either could not be located or declined to participate. Each participating farmer was offered a

[2] Before the beginning of this study, some biochar dissemination work was performed in the area, by a private business (re:char) and an NGO (ACON). To avoid any residual influence from those organizations, areas around Bungoma were selected that were at least 4km away from areas where either organization had worked.

[3] Families in this region are commonly polygamous, though some polygamous families farm separately, while others farm collectively. Thus, a household could be a monogamous family, a polygamous family farming together, or one of several wives of a single husband who farm separately, together with the husband or separate from the husband if he farms separately from his wives. A person "farmed separately" if they were the sole agricultural decision maker for a particular plot of land.

[4] Assuming a maize price of \$392/ton (the approximate average retail price during the project), expected yield increases are sufficient to render biochar profitable in expectation at a cost up to \$121/ha (KSh1038/quarter acre), all else being equal. This is less than the cost of the most-expensive randomly-allocated price of KSh1,250 (\$14.7) per one-quarter acre, which is equivalent to \$145.3/ha.

[5] Most farmers in the sample consume the majority of the maize that they produce; very few sell in denominations larger than the 90kg sack. We therefore use prevailing retail prices during the pre-harvest season (when most farmers stores have been depleted) for 90kg sacks, rather than bulk grain prices. See Burke (2014) for a discussion of maize storage among smallholder farmers in the Bungoma region.

Table 1. Summary Statistics Describing the Sample

Statistic	N	Mean	St. Dev.	Min.	Max.
Maize Yields (tons/ha)	864	1.093	0.81	0	3.34
Hectares Maize	864	0.35	0.320	0.05	3.24
Hectares Sugar	889	0.1	0.262	0	3.64
Total Land (ha)	889	0.56	1.752	0	51
Fertilizer Expenditure ($/ha)	887	78.39	72.063	0	374
Income (Sep 2014)($/month)	895	61.79	108.32	0	1,458
Family size	911	5.89	2.49	1	16
Member of One Acre Fund	914	0.23	–	0	1
Gender main farmer (0=female)	914	0.35	–	0	1
Ever produced charcoal	914	0.33	–	0	1

Note: Time-varying agricultural variables (yields, land allocation, etc.) taken from the long rains 2013, before any of the treatments were implemented.

selection of sealed, unmarked envelopes, of which they took one. Each envelope contained a price (each farmer saw only one price), and some envelopes contained vouchers for "risk-free trial" (RFT) offers. Under the RFT, farmers could get biochar, but not be obligated to pay until after the subsequent harvest. If at that point they did not feel that the result was sufficient to justify the expense, they were not obligated to pay, though they would be excluded from future sales. This distribution of prices and RFT offers is given in appendix A.2.

Accepting the offer consisted of having our team visit their farm and apply biochar to one-quarter acre (.1 ha). Delivery and application were included in the price in order to abstract from issues of compliance, skill, and transport costs. The application rate (0.5 Mg/ha) was controlled to remain roughly uniform across all who received biochar.

Subsequent uptake was very low absent large inducements. Only 18% of farmers took biochar, and of those, 90% had RFTs. Only 2.6% of farmers without RFTs purchased biochar, and those who did had an average subsidy of 77%. Biochar adoption by individuals was not significantly related to the share of their network with demonstration plots. More detail is given in appendix A.3.

Second sales offer. While very few farmers purchased biochar in the previous season, the large number of RFTs ensured that a large number of examples existed to demonstrate biochar's performance. On average, yields on biochar-amended portions of farms were much higher than on portions left unamended (figure A.3). As with the previous season, yield differences are larger on fields where unamended yields were low. The project began to receive inquiries about biochar sales for the coming short rains season well before the long rains harvest was in, indicating that biochar's impact was readily apparent among the farmers in the study, and to their social contacts.

Project staff weighed the harvests for all farmers who had gotten biochar for the 2014 long rains (in early 2014), from portions with and without biochar (figure A.3). Biochar plots substantially out-performed plots without biochar, and again on the fields of farmers whose yields had previously been low. These yield differences—as observed through our estimated network by individuals—are those that form $\bar{\Delta}Y_{\boldsymbol{A}_i}$. Note that no attempt was made to enforce equality of input use between plots in this season—this was felt to be inappropriate given that farmers were paying for biochar with their own money.

Because sales for the 2014 long rains season were so low absent the RFT, prices were reduced and more farmers were offered higher subsidies. In addition, all farmers who had gotten a RFT in the previous season were offered a randomized discount on the cost of their RFT repayment. Of the 161 respondents with RFTs, only 19 repaid ahead of the 2014 short rains planting season.[6] Between August and October 2014, 96 farmers bought biochar, including 16 of those who

[6] Propensity to repay was negatively correlated with repayment price and with previous-month income, but was uncorrelated with measured results on demonstration plots, and uncorrelated with reported impressions of outcomes on demonstration plots (table A.1). It is possible that low repayment stems from the objective assessment by farmers that biochar was unprofitable to use. Alternatively, it could be the case that paying for the RFT is behaviourally difficult, because it involves incurring a present cost for a past benefit. Focus group discussions support the view that RFT repayment was perceived simply as an addition to the price faced in the current season.

chose to repay their RFT. The average price at which biochar was purchased was KSh143 (USD1.7) per 25kg sack, similar to the average price of KSh294 (USD3.5) in the previous season for 50kg. More detail is given in appendix A.3.

Randomization Checks, Attrition and Compliance

Randomized treatments are approximately orthogonal to observable covariates (table A.2). However, the study faced substantial nonresponse throughout the course of several survey rounds. Most attrition stemmed from the inability of enumerators to locate respondents. This owed largely to respondents moving (often temporarily) for off-farm work. From an initial census of 1,115, we possess complete observations of all variables in our empirical model for 823 farmers.

While nonresponse is approximately uncorrelated with randomized treatments, 7 of the 75 farmers who were offered demonstration plots did not take them—largely because they did not plan to farm in the subsequent season. This creates the possibility that network injection points were nonrandom. We address this possibility by creating a propensity score for receiving a demonstration plot, and condition our estimates on it throughout. This is described in more detail in appendix A.2.

Methods

This section describes social network estimation and imputation methods used, the econometric methods used, and the empirical strategy.

Social Networks

A social network survey was conducted in October and November 2013. Each respondent was asked about 75 other respondents from throughout the four zones, with the probability that farmer i was asked about farmer j being inversely proportional to the physical distance between them. For each potential link, farmer i was asked if they knew farmer j, using nicknames, spouse names, and place names to help distinguish between people with similar names. Where farmer i indicated that they knew j, surveyors asked about a number of different sorts of linkages. Importantly for this article, any farmer i who knew j was asked if they regularly speak with j about farming—we use this to construct our network of farmers who discuss farming.

The average farmer knew about one-third of the 75 that were asked about. It is not assumed that a farmer who says that they know another farmer is known by that farmer—a directed network is assumed. While the literature is dominated by analysis of undirected networks, assuming an undirected network is problematic in this context—of the 7,953 instances where both farmer i was asked about farmer j and vice versa, the link was reciprocal only 67.6% of the time.

Inferences based on sampled networks are generally biased (Chandrasekhar and Lewis 2011). While we possess a census of $N = 945$ nodes in our network, only $75N$ of the $N(N-1)$ links are observed through the survey. Resultant bias is avoided by imputing missing edges with their conditional expectations, which are estimated using a random forest (Breiman 2001).

Random forests are a nonparametric classification and regression algorithm. Like most methods from the machine learning literature (and unlike much of classical econometrics, especially microeconometrics), random forests are statistical tools developed for prediction, rather than parameter estimation (Mullainathan and Spiess 2017). As such, they provide excellent out-of-sample predictive performance for many problems, at the expense of parsimonious interpretability or clear linkage with any underlying theory about the data-generating process. Random forests work by building an ensemble of decision trees, using bootstrapped samples of the data for each tree and random subsets of the explanatory variables at each split in the tree. See Friedman, Hastie, and Tibshirani (2001) for an overview. We fit separate random forests to predict general linkage, and linkage with respect to talking about farming.

Variables used to train the random forest, along with metrics of variable importance and predictive accuracy, are given in figure 1. Of the variables, distance has the most predictive power, followed by the proportion of people that a given farmer knows and is reported as being known by. Other variables are less predictive. Estimated out-of-sample misclassification error is 15.7%, while error in probability (i.e., $\mathbb{E}_{OOS}|\text{link} - \hat{pr}(\text{link})|$) is 22%. The forest is used to impute links with their conditional expectation (i.e., estimated probabilities of linkage, rather than most likely classes) where missing.

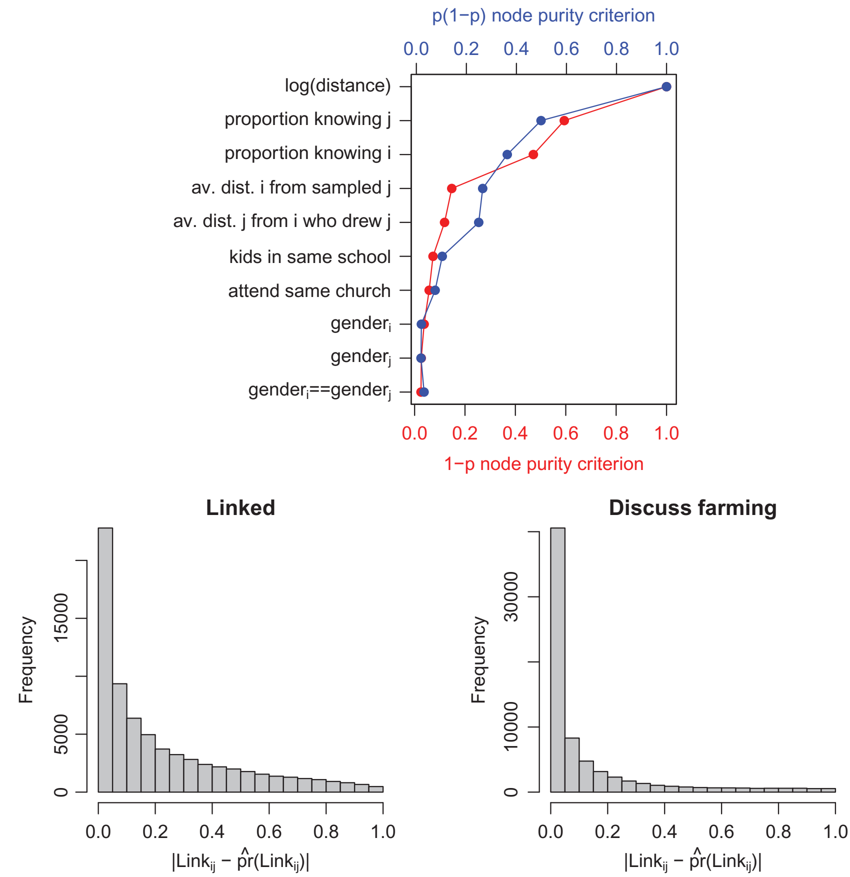

Figure 1. Results of a random forest classifier

Note: Used to impute social network links with conditional expectations where missing. Variable *i* refers to farmers who were asked about fellow farmers, *j*. The upper plot gives relative measures of the importance of variables, based on measures of node purity (i.e., the ability of a given variable to correctly classify observations that were not part of a given tree's bootstrap sample, where *p* is the proportion correctly classified—see Friedman, Hastie, and Tibshirani (2001) for a detailed description of these criteria. The histograms give the distribution of difference between linkage and estimated probability of linkage.

Social network statistics are given in table 2. In our estimated network, the average farmer knew 106 people, and spoke about farming with about one-quarter of them. Along other dimensions, statistics measured along the full networks are comparable to those in the subnetworks; on average, shares with various treatments, moments of network yields, and moments of response to biochar are comparable.

Social network data was collected after the demonstration plots were established, and only collected once. Our interpretation of subsequent empirical results relies on the assumption that actions taken by the project did not alter social network composition.

Sub-networks. Our conceptual model implies that the relationship between adoption and unconditional moments of adopters' outcomes will be weaker in networks that exchange a lot of information, and stronger in networks that exchange less information. Given a single social network adjacency matrix $\boldsymbol{A}$, representing a general sort of linkage (say, farmers who are acquainted with one another), one may construct a subnetwork by zeroing all elements that do not represent a particular sort of linkage.

Given data on which links convey information about farming, sub-networks may be constructed that represent high or low levels of information transfer with respect to farming. Specifically, the former would transmit signals of the triple $[\Delta Y_j, B_j, \boldsymbol{X}_j]$ to a receiving farmer *i*, while the latter would not contain information about $\boldsymbol{X}_j$. While we do not observe the flow of information about farming, we do know who reports taking about farming with who. From this proxy, we construct a (partially-imputed)

Table 2. Summary Social Network Statistics, Including all Explanatory and Instrumental Variables

Description	Name	N	Mean	St. Dev.	Min.	Max.
N. Links (of 75)		945	26.3	10.8	0	61
+*talkfarm*		945	7.2	8.4	0	55
-*talkfarm*		945	19.2	10.7	0	52
Estimated N. Links (of 945)		945	106.4	54.5	19.1	354.3
+*talkfarm*		945	26.0	24.4	2.1	354.0
-*talkfarm*		945	82.7	41.5	12.9	273.9
Share with demo (SR13)		945	0.06	0.02	0.01	0.14
+*talkfarm*	$S^{SR13}_{\mathbf{A}_i}$	945	0.06	0.04	0.005	0.29
-*talkfarm*		945	0.063	0.021	0.013	0.15
Share adopting (LR14)		945	0.13	0.03	0.05	0.32
+*talkfarm*	$S^{LR14}_{\mathbf{A}_i}$	945	0.13	0.058	0.025	0.51
-*talkfarm*		945	0.13	0.030	0.054	0.25
Share with RFT (LR14)		945	0.12	0.03	0.04	0.307
+*talkfarm*	$S^{RFT}_{\mathbf{A}_i}$	945	0.12	0.054	0.013	0.47
-*talkfarm*		945	0.12	0.029	0.050	0.25
Share with low price (LR14)		945	0.064	0.03	0.01	0.15
+*talkfarm*	$S^{\leq 250}_{\mathbf{A}_i}$	945	0.063	0.038	0.005	0.29
-*talkfarm*		945	0.064	0.026	0.013	0.16
Average network yields		945	5.9	0.52	4.9	7.8
+*talkfarm*	$\bar{Y}^{T<t}_{\mathbf{A}_i}$	945	6.049	0.67	3.39	9.27
-*talkfarm*		945	5.93	0.52	4.81	7.68
SD network yields		945	3.2	0.206	2.675	3.9
+*talkfarm*	$s(Y^{T<t}_{\mathbf{A}_i})$	945	3.24	0.32	2.23	4.50
-*talkfarm*		945	3.22	0.20	2.78	4.08
Average network Δyields		945	0.83	0.36	–0.51	3.19
+*talkfarm*	$\bar{\Delta} Y_{\mathbf{A}_i}$	945	0.87	0.60	–1.98	6.15
-*talkfarm*		945	0.83	0.35	–0.38	3.94
SD network Δyields		945	1.5	0.44	0.707	3.75
+*talkfarm*	$\sigma^{U}_{\mathbf{A}_i}$	945	1.52	0.55	0.50	5.18
-*talkfarm*		945	1.50	0.43	0.69	3.96
Prev. yields × in-network $\hat{pr}$(adopt)		945	5.94	0.67	4.44	8.40
+*talkfarm*	$\bar{Y}^{\hat{pr}(BC)}_{\mathbf{A}_i T<t}$	945	6.05	0.94	2.98	10.67
-*talkfarm*		945	5.92	0.61	4.45	8.22
SD prev. yields × in-network $\hat{pr}$(adopt)		945	3.16	0.29	2.35	4.59
+*talkfarm*	$s(Y_{\mathbf{A}_i})^{\hat{pr}(BC)}$	945	3.18	0.44	1.75	5.19
-*talkfarm*		945	3.16	0.30	2.36	4.76

Note: Statistics of yield change outcomes are reported in units of 90kg bags/acre (convertible to Mg/ha by multiplying by ~0.22), and reflect the distributions of those statistics as observed by nodes in the network, not the true distributions, which are given in figure A.3.

network representing farmers who do or do not share agricultural information.

Empirical Strategy. The empirical analog to our conceptual model is

$$\text{(11)}\quad \text{Adopt}_i = \boldsymbol{1}\Big(\alpha + \beta_1(\text{adopt}^{LR14}) + f_1(S^{LR14}) + f_2(\bar{\Delta} Y_{\mathbf{A}_i}) + f_3(\sigma^{U}_{\mathbf{A}_i}) + \sum_{j=4:J} f_j(Z_j) + \epsilon > 0\Big).$$

This is the semiparametric representaiton of equation (10), with the inclusion of an effect of prior adoption (i.e., in the 2014 long rains season), as well as the share of the network that had previously adopted. Further, $\boldsymbol{Z}$ includes biochar price—which appears in our conceptual model as p_B—as well as several controls: distance from the center of the project zone, the share of contacts having had a demonstration plot, the mean and variance of pre-project maize yields in the network, demonstration plot propensity score and its average value in the network, and dummies for having had a demonstration plot, gender, past charcoal production, familiarity with biochar-disseminating organizations, and location. We ignore the ancillary costs of adoption, p_Δ, which are close to zero in our context given (*a*) the small-scale of the biochar applications, (*b*) the fact that the project hired workers to apply the biochar to the fields of farmers, and (*c*) the fact that biochar does not specifically require complementary inputs.

Note that equation (11) does not include terms d_μ or d_σ representing the difference between the conditional and unconditional moments—these are unobserved, and are only zero when farmers observe none of the mediators of the results that they observe.

Instruments for endogenous variables. The first four terms of model (11) are endogenous. Instruments for them are functions of randomized inducements to adopt, interacted with characteristics of farmers as observed by individuals through the social network.

Biochar adoption in the previous season is instrumented with the (randomized) price at which biochar was offered, as well as an interaction between the price and the risk-free trial offer[7].

The share (or proportion) of the social network adopting biochar, $S^{RFT}_{\mathbf{A}_i}$, is instrumented with a nonparametric interaction of $S^{RFT}_{\mathbf{A}_i}$, the share of the network receiving the risk-free trial offer, and $S^{\leq 250}_{\mathbf{A}_i}$, the share of the social network receiving a low price (below 250KSh, or \$2.90). These proportions are clearly exogenous, given that the network was randomly shocked.

Our instrument for unconditional mean yield change, $\bar{\Delta}Y_{\mathbf{A}_i}$, is motivated by the fact that biochar's benefit is negatively correlated with previous yields—its benefits are larger in poor soils. Figure A.3 gives evidence for this—responses to biochar in both seasons was greater when yields on unamended portions were low. Average benefits in the network are thus a function of *which* network members adopted biochar—an individual whose adopting network contacts tended to get low yields would see larger yield changes than someone whose adopting network contacts tended to get better yields. If adoption were exogenous, we would thus use the average pre-project yields of those in the network that had adopted.

Adoption is of course endogenous, so we use purely exogenous treatments (price, demonstration plots, and RFTs) to form a propensity score for adoption. The instrument is the average pre-project yield in the network, weighted by this propensity—which is exogenous, being a function of exogenous variables; it is denoted $\bar{Y}^{\widehat{pr}(BC)}_{\mathbf{A}_i T<t}$.

This instrument will fail the exclusion restriction inasmuch as farmers with higher yields among known adopters will likely have higher yields in their network more generally. This in turn may be correlated with characteristics of the farmer that determine adoption. We sidestep this issue by explicitly controlling for average network yields $\bar{Y}^{T<t}_{\mathbf{A}_i}$ in the first and second stages.

Our instrument for the unconditional standard deviation of yield change in the network, $\sigma^U_{\mathbf{A}_i}$, is analogous. We control for the standard deviation of pre-project maize yields weighted by the exogenous adoption propensity, and control for the standard deviation of pre-project maize yields in the network.

Econometric Methods

Two-stage least squares estimates of equation (11) will give biased marginal effects because there are multiple overlapping treatments in a model with a binary outcome. If a discrete outcome y is determined by whether a latent parameter crosses a threshold $y = \mathbf{1}(\boldsymbol{X\beta} + \epsilon > 0)$, then the marginal effect of changing any variable $x_p \in \boldsymbol{X}$ is $\frac{\partial pr(y)}{\partial x_p} = \lambda(\boldsymbol{X\beta})\beta_p$, where Λ is a link function and λ is its derivative, the form of which corresponds to the density of ϵ in the latent variable. This marginal effect depends on covariates $\boldsymbol{X}$; the effect of changing any given variable depends on the values of the others, as captured by the derivative of the link function. If ϵ is continuous and uncorrelated with $\boldsymbol{X}$, there is no function Λ for which $\frac{\partial pr(y)}{\partial x_p} = \beta_p$.

If there are two treatments, D_1 and $D_2 \in \boldsymbol{X}$, then calculation of the marginal effect of either necessarily involves specifying a value for the other. If interest centers on the population from which the experiment sampled, and that population is untreated, then the marginal effect of D_1 should be calculated at $D_2 = 0$. Given that the partial derivative of a LPM with respect to either treatment is invariant to the other, marginal effect estimates from linear probability models will necessarily be biased away from the population effect in these circumstances, unless the effect of the other

[7] The main effect of the risk-free trial on adoption is not an instrument, as it directly influences adoption in the 2014 short rains season, thereby failing the exclusion restriction.

treatment is exactly zero. Naturally this problem is no longer relevant where there is only one treatment, as the covariates of the sample are similar to those of the population.

Moving away from the LPM presents challenges, however; standard instrumental variables methods do not transfer directly to limited dependent variable models. We will therefore consider two methods in turn for estimating equation (11)—control functions and special regressors.

Control functions. Dating at least from Heckman (1978), the control functions (CF) approach adjusts for endogeneity in variables $\boldsymbol{X}^{en}$ by controlling for the residuals of first-stage regressions $x_p = f(\boldsymbol{X}^{ex}, \boldsymbol{I}) + \zeta$, where p indexes endogenous variables, and $\boldsymbol{I}$ are the instruments. In the second-stage regressions, one can fit $g(y) = \sum_j f_j(X_j^{ex}) + \sum_k f_k(X_k^{en}) + \sum_p f_p(\hat{\zeta}_p) + \epsilon$, where ζ_p is a vector of residuals from the pth first-stage regression. When $g(y) = y$, CF produces identical estimates to standard linear instrumental variables methods; its key advantage is consistency when $g(y)$ is nonlinear, such as in probit.

The control functions approach has important limitations when used with nonlinear link functions. First, CF is inconsistent when endogenous variables are not continuously distributed. In our context, all endogenous variables are continuous (or can be made so by taking logs), except for the decision to purchase biochar in early 2014. This violation of CF assumptions is one motivation for the use of special regressors (described in the next section).

More importantly, CF is not robust to misspecification: the omission of relevant instruments will yield a $\hat{\zeta}$ that does not fully address the endogeneity problem when controlled for in the second stage. This requirement is fundamentally untestable, likely violated in our context, and is another motivation for special regressor estimation.

While CF can yield consistent estimates of index coefficients when its conditions for consistency are met, interest typically centers on the average structural function (ASF; Blundell and Powell 2003, 2004), which measures the change in the response variable with changes in explanatory variables. The ASF is calculated by first estimating the distribution of the latent error $F_{-\epsilon} = \mathbb{E}[y|\boldsymbol{X}\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\zeta}}_{\boldsymbol{i}}]$, and then taking predicted values of the index function across a grid of a particular variable, with covariates held constant and $\hat{\zeta}$ averaged out, transformed according to $\hat{F}_{-\epsilon}$. Often in practice, however, $F_{-\epsilon}$ is assumed to be normal.

The CF estimators are simple to compute with standard statistical software. We compute an asymptotically valid covariance matrix – necessary to adjust for the fact that $\boldsymbol{\zeta}$ is estimated from data – via a parametric bootstrap, following Marra and Radice (2011).[8]

Special regressors. When CF assumptions are not met, a more robust approach to binary choice models with endogeneity is the “special regressors” (SR) approach of Lewbel (2000). This approach recovers the flexibility of linear 2SLS, including consistency in the presence of non-continuous instruments and incomplete subsets of instruments. This is in contrast to CF, which is not robust to these things.

The SR method exploits a continuous exogenous regressor V within a model $y = \boldsymbol{I}(\boldsymbol{X\beta} + V + \epsilon > 0)$. It is motivated by the fact that the model can be recast as the cumulative density function with the mean vector $\boldsymbol{X\beta}$ evaluated at $v : \mathbb{E}(y|V = v) = pr(X\beta \geq -v) = 1 - pr(X\beta < -v) = 1 - F_{X\beta}(-v)$. If $\hat{H}(v)$ is a consistent nonparametric estimator of $\mathbb{E}(y|V = v)$, then $\hat{H}(X\beta)$ is a consistent estimator of the distribution of the mean vector. The actual estimator builds on the above logic to construct a continuous measure $T = (y - (V > 0))/f(U)$, where $f(U)$ is the conditional probability density of the special regressor given the data, and the special regressor V is centered on zero and positively correlated with the outcome. Dong and Lewbel (2015) show that $\mathbb{E}(T) = \mathbb{E}(\boldsymbol{X\beta})$, such that a linear 2SLS regression on $\hat{T}$ instead of the original binary response yields a consistent estimate of $\boldsymbol{\beta}$.

Beyond exogeneity and continuousness, a valid special regressor requires large support, in the sense that the distribution of $\boldsymbol{X\beta} + \epsilon$ is contained within the distribution of – V. See Lewbel et al. (2012b) for an exposition of the estimator and the derivation of the support requirements. For the short rains 2014 adoption season, the only available special regressor will be randomized biochar price, residualized via a regression on all available covariates to make it continuous.[9]

[8] A nonparametric bootstrap is feasible, but much more computationally expensive in larger datasets where asymptotic normality is more relevant.

[9] For example, if $\boldsymbol{X}$ are exogenous covariates, then price is residualized by estimating $price = \boldsymbol{X\beta} + \epsilon$, and saving $price - \widehat{price}$.

Lewbel, Dong, and Yang (2012a) propose estimating marginal effects by first fitting a nonparametric regression of the outcome on the index $y = h(\boldsymbol{X}\hat{\boldsymbol{\beta}} + V)$. Marginal effects are then estimated by taking the fitted values across a grid of $x \in \boldsymbol{X}$, or V. This is termed the "average index function" (AIF).

Generalized additive models. To avoid both bias and inefficiency stemming from functional form mis-specification (as distinct from omitted control functions), this article extensively employs generalized additive models. Generalized additive models (GAMs; Hastie and Tibshirani 1990) are semiparametric generalizations of generalized linear models (GLMs). The general form is $g(\mu) = \boldsymbol{W}'\boldsymbol{\theta} + \mathcal{F}(\boldsymbol{X}) + \epsilon$, where μ is the mean of y, $\boldsymbol{W}$ are variables associated with linear slope coefficients, and $\boldsymbol{X}$ is a matrix of variables represented nonparametrically. The function $\mathcal{F}(.)$ is commonly a sum of univariate smooth terms, but can also include smooth functions of more than one variable (Wood 2006b). These smooth functions are typically represented by penalized splines; specifying $\mathcal{F}(.)$ terms using some spline basis b such that $f_p(x) = \sum_k^K \gamma_k b_k(x)$. We use thin-plate regression splines (Wood 2003) throughout, which are less sensitive to knot location than many alternatives; k is typically chosen with a higher basis dimension than is needed (We use 10 per variable)—and $\hat{\boldsymbol{\beta}}$ is chosen to minimize $-l(\boldsymbol{\beta}) + \sum_{i=1}^m \lambda_i \boldsymbol{\beta}' \boldsymbol{S}_i \boldsymbol{\beta}$, where $l(\boldsymbol{\beta})$ is the log likelihood of the model, m is the number of nonparametric terms, $\boldsymbol{\beta}$ is defined to include both $\boldsymbol{\theta}$ and the penalized coefficients γ associated with each spline term. The smoothing penalties λ_i control overfitting, which is common to un-penalized spline models, and the matrices $\boldsymbol{S}_i$ are constructed so that smoothing only applies to nonparametrically-represented terms.

The λ_i are chosen to minimize bias while avoiding overfitting. This article uses the restricted maximum likelihood smoothing parameter selection criterion (Wood 2011) throughout, which is similar to the treatment of smoothing parameters as shrinkage factors in random effects models.[10] When $\lambda \to \infty$, the result is equivalent to a generalized linear model (or OLS where y is continuous); GAMs include these as a special case. Where λ is constrained to be 0, the result is an unpenalized spline model. These will typically be overfit, particularly where basis dimension is sufficiently high to avoid substantial bias. Where λ is estimated from the data however, the result is a flexible model in which both the functional form and the degree of smoothing are data-driven.

While less flexible than fully nonparametric techniques (that do not require additivity assumptions), GAMs can be a clear improvement over parametric techniques when continuous variables have nonlinear effects. They are particularly convenient in the present context, where the assumed mean-variance utility implies additivity, but not linearity in additive terms. Where effects of additive terms truly are linear, smoothing parameter selection algorithms will typically recover this. Textbooks on the penalized regression approach to GAMs include (Ruppert, Wand, and Carroll 2003, and Wood 2006a).[11]

The methodological contribution proposed in this article is twofold. First, we extend the SR framework to admit GAMs. The extension to CF estimators is straightforward, and has been proposed by Marra and Radice (2011). The SR extension is also straightforward. Quite simply, all OLS regressions in the SR algorithm are replaced with (Gaussian) additive models (with smoothing penalties chosen as described above). In particular, the SR-GAM approach replaces a 2SLS regression of $\hat{T}$ on $\boldsymbol{X}$ (using instruments $\boldsymbol{I}$) with a CF estimator:

$$(12) \quad \hat{T} = \sum_j f_j(X_j^{ex}) + \sum_k f_k(X_k^{en}) + \sum_p f_p(\tilde{\zeta}_p) + \epsilon.$$

Note that $\boldsymbol{X}^{ex}$ does not include the special regressor V, and that $\tilde{\zeta}$ replaces $\hat{\zeta}$. We define $\tilde{\zeta}$ as the matrix of control function residuals derived from first-stage regressions using whatever instruments are available, rather than a "complete" set of instruments required for

[10] Other approaches exist, including generalized cross-validation—these methods choose λ in order to minimize prediction error, but are sometimes prone to undersmoothing.

[11] Books by Li and Racine (2007), and Friedman, Hastie, and Tibshirani (2001) describe other approaches to estimating GAMs—by marginal integration and backfitting algorithms, respectively.

consistency under the control functions when outcomes are binary. Because $\hat{T}$ is continuous and control functions yield identical results to 2SLS in linear[12] models like equation (12), estimates of the coefficients comprising $f()$ terms are consistent conditional on the available instruments being valid.

Because $\tilde{\zeta}$ is estimated from data, however, inference based on equation (12) may be misleading. As described above for the CF estimator, we use a parametric bootstrap to adjust the variance-covariance matrix to account for uncertainty in $\tilde{\zeta}$.

Refinements to special regressor marginal effects. The second methodological contribution of this article is the introduction of the *estimable ASF estimator*, which can alternatively be considered an *augmented AIF estimator*. The average structural function of Blundell and Powell (2004) is based on the cumulative distribution function of the latent error of an index model, denoted as $F_{-\epsilon}$. It is estimated as the conditional expectation of the outcome given the index $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ and control function residuals $\hat{\zeta}$. If $\hat{\zeta}$ is complete, then $\mathbb{E}[y|\boldsymbol{X}\hat{\boldsymbol{\beta}}, \hat{\zeta}] = F_{-\epsilon}$ can be consistently estimated by nonparametric regression, and marginal effects are computed as $\hat{\beta}(N^{-1}\sum_i \hat{f}_\epsilon(\boldsymbol{X}\hat{\boldsymbol{\beta}}, \hat{\zeta}_{\boldsymbol{i}}))$. Completeness in $\hat{\zeta}$ may be an unrealistic assumption in many contexts, however, including the present context.

The average index function of Lewbel, Dong, and Yang (2012a) is essentially the same thing, but does not condition on $\hat{\zeta}$, which is not a product of Lewbel, Dong, and Yang's framework. Rather, it is calculated as $F_{-\epsilon|\boldsymbol{X}\hat{\boldsymbol{\beta}}} = \mathbb{E}[y|\boldsymbol{X}\hat{\boldsymbol{\beta}}]$, and marginal effects are $\hat{\beta}(f_{\epsilon|\boldsymbol{X}\hat{\boldsymbol{\beta}}}(\boldsymbol{X}\hat{\boldsymbol{\beta}}))$. Lin and Wooldridge (2015) point out that the regression $y = H(\boldsymbol{X}\hat{\boldsymbol{\beta}}) + \epsilon$ suffers from correlation between the ϵ and $\boldsymbol{X}$, which can lead to incorrect marginal effects in many cases. However, their proposed solution—assuming that the error is normally distributed—may also be wrong in practice, and lead to incorrect marginal effects (though they will have the same sign as $\hat{\beta}$).

The approach taken here is to simply swap $\hat{\zeta}$ for $\tilde{\zeta}$ in calculating the ASF. This approach does not eliminate bias in the estimation of $\hat{F}_{-\epsilon}$, and is inconsistent: the estimator for $\tilde{F}_{-\epsilon}$ used to calculate the estimable ASF can be viewed as an error-in-variables model with $\tilde{\zeta} = \hat{\zeta} + \boldsymbol{\eta}$. As such, the effect of the control functions will be attenuated, and the estimate of the effect of the index on the outcome will be rendered inconsistent. The sign of this inconsistency will depend on exactly which instruments are omitted, and what effect they have on the endogenous variables. Despite these shortcomings, this approach may be better than a normal approximation to the distribution of ϵ—though monotonicity of $\tilde{F}_{-\epsilon}$ should be checked to ensure that marginal effects have the correct sign. This approach uses more information than the AIF as originally proposed by Lewbel, Dong, and Yang (2012a)—which can be considered to assume that $\tilde{\zeta} = 0$—but nonetheless is not guaranteed to give a better approximation of $F_{-\epsilon}$. The difference between the two will depend on the degree to which $\tilde{\zeta}$ is a good proxy for $\hat{\zeta}$. Results are reported based both on the estimable ASF, the AIF, and an assumption of normality in ϵ.

Results

This section fits our empirical model to data from the short-rains 2014 adoption season.[13] We begin with data derived using the "full" network, in which a link is defined simply by *i* knowing *j*. We then fit the same model to data derived from subnetworks, in which links are defined by *i* knowing *j* and reporting whether or not the two discuss farming.

Full Network

Marginal effect estimates are given in figure 2. The full output of the underlying models is in figure A7. Models all control for zone dummies, gender, prior experience either with biochar or the organizations that had promoted it, the physical distance of *i* from the center of their zone, and demonstration plot propensity scores (see discussion of attrition, above).

[12] To be clear, the underlying representation of equation (12) is linear in the parameters of the spline basis expansion.

[13] Data from the long rains 2014 season is not used because adoption in that season was entirely dominated by the risk-free trial. By bounding potential outcomes at zero, the RFT may alter a farmer's implicit estimate of the mean and variance of adoption. Thus, our conceptual model would imply that the RFT would effectively decorrelate adoption from $\bar{\Delta}\boldsymbol{Y}_{\boldsymbol{A}_i}$ and $var(\Delta\boldsymbol{Y}_{\boldsymbol{A}_i})$. A test of the model would require subsetting the data to those who did not receive the RFT. Of those, only 16 adopted biochar, which is 2.5% of the sample.

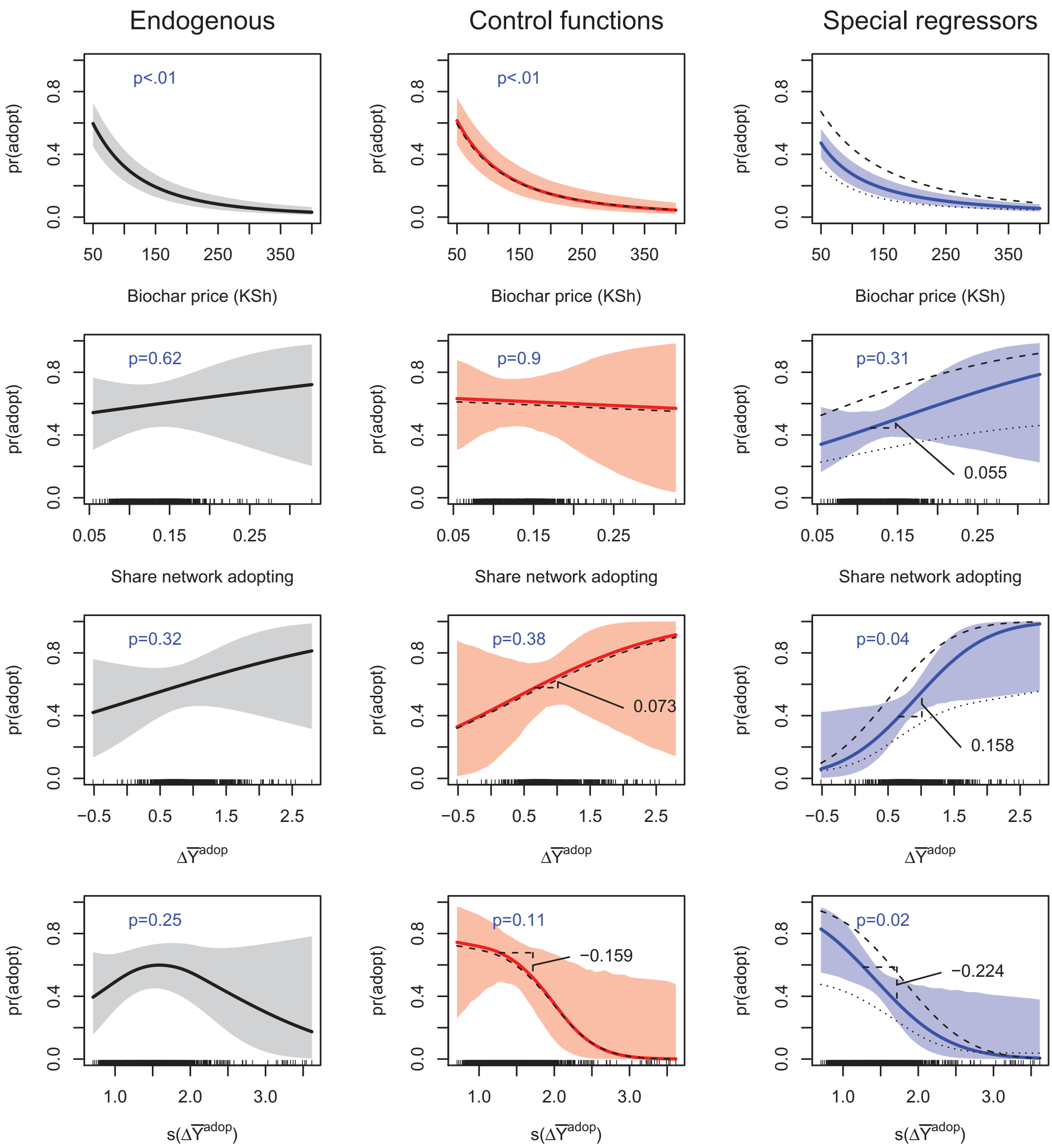

Figure 2. Main endogenous, control functions, and special regressors estimates

Note: Full results (complete output of endogenous models, first and second stages, and SR index functions) are given in figure A.7. Both control functions and special regressor estimates give estimable average structural functions. The ASF/AIF estimates hold all covariates to their means, all treatments to zero, but biochar price at the minimum price offered (KSh 50 [$0.6] per 25kg sack). Shaded regions give 95% confidence intervals. Printed p-values based on reduced-rank Wald tests that the penalized spline terms in the underlying GAM are jointly different from **0** (Wood 2013). Printed average marginal effects shown over the interval $\mathbb{E}(x)\pm sd(x)/2$. Dotted lines give normal approximations to $F_{-\epsilon}$, while dashed lines give AIF estimates unconditional on $\tilde{\zeta}$.

Farmers who observed better average response to biochar in their networks were more likely to adopt. The SR (CF) estimator suggests an average marginal effect of 15.8 (7.3) percentage points. The CF estimate is not significant, at $p \approx .38$.

Farmers who observed variable response to biochar in their networks were less likely to adopt. The SR (CF) estimator suggests an average marginal effect of –22.4 (–15.9) percentage points. These results are consistent with substantial risk aversion.

In all cases, the eASF estimate is larger than the AIF (dotted line), and gives steeper marginal effects. Likewise, the eASF suggests a smaller marginal effect than does a normal

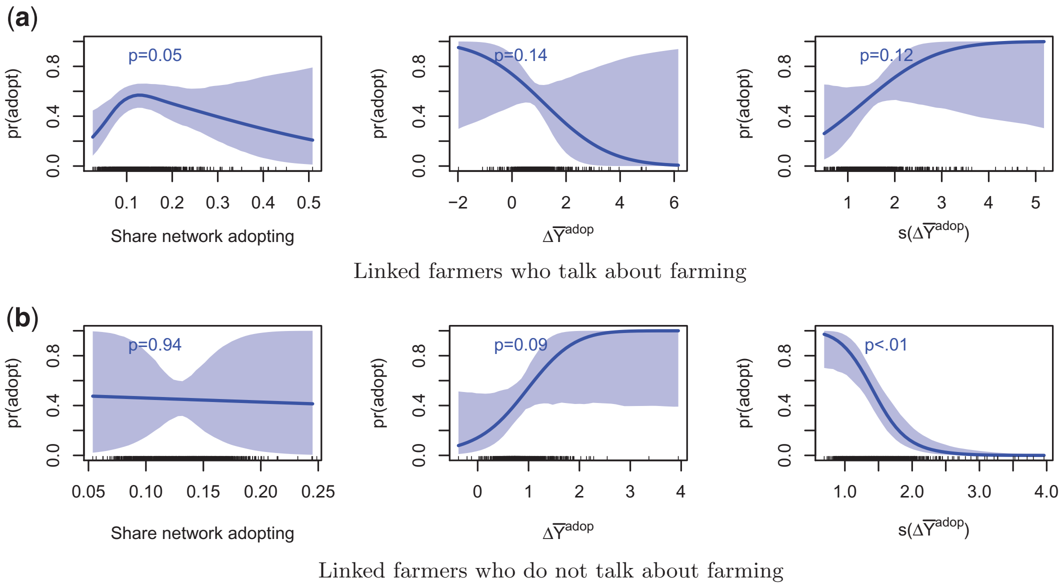

Figure 3. Marginal effects of model 11, fit to sub-networks

Note: Shaded regions give 95% confidence intervals. Printed p-values based on reduced-rank Wald tests that the penalized spline terms in the underlying GAM are jointly different from **0** (Wood 2013).

approximation (dashed line) to the latent error distribution $F_{-\epsilon}$.

Column 1 of figure 2 shows estimates that do not account for endogeneity in observed response to biochar. The estimation results are insignificant, and the direction of their difference from the SR and CF estimates is consistent with the hypothesized direction of bias due to endogeneity from our conceptual model.

The SR estimates are probably more reliable than CF estimates; CF consistency requires that the instrumental variables entirely solve the endogeneity problem through the creation of auxiliary variables. There are probably many factors that both affect adoption and are correlated with moments of network yield, and that are left unmodeled here. One example could be localized pest outbreaks: these may affect both a farmer's propensity to adopt biochar, as well as the distribution of biochar results that they have observed.

On the other hand, SR consistency requires that the special regressor has large support with respect to the distribution of the latent variable. This does hold—the standard deviations of the special regressor and the estimated index are 0.54 vs. 0.51, respectively. However, the distribution of the estimated index is not entirely contained within that of the special regressor (see figure A7c). This means that the SR does not take on values that cause adoption with 100% certainty. In other words, we did not offer biochar at sufficiently low prices to drive non-adoption to zero. As such, $F_{-\epsilon}$ is extrapolated beyond about 95% probability of adoption.[14] The assumption that the function continues to asymptotically approach 1 appears innocuous.

Sub-network Results

Our conceptual model predicts that adoption will be more closely related to unconditional moments in the network when farmers share little information. This is because they have less information with which to constrain a conditional mean and variance.

On the other hand, our model predicts that farmers who share information will not be as

[14] This extrapolation is straightforward. The index function $E[y|\boldsymbol{X}\hat{\boldsymbol{\beta}}+V,\tilde{\zeta}]$ is approximated with $g(y)=\alpha+f_1(\boldsymbol{X}\hat{\boldsymbol{\beta}}+V)+\sum_p f_p(\tilde{\zeta}_p)+u$, which is estimated by a probit-linked GAM. The smoothing parameter selection algorithm penalizes the coefficients comprising $f_1()$ down quite closely to the equivalent of a linear probit fit. Extrapolation amounts to extending this probit beyond the range of the observed index, and a parametric bootstrap applied to the coefficients of this penalized probit allows confidence intervals to be drawn around these extrapolations, which widen beyond the largest observed values of the estimated index.

sensitive to unconditional moments. This is because they will use the information that they collect through their network to construct models explaining the results that they observe, and the conditional means and variances derived from these models will generally be different from their unconditional counterparts.

Results (3) are consistent with the conceptual model. Farmers who say that they speak to one another about farming are not sensitive to the unconditional moments of the outcomes within their networks, but they are sensitive the share of their network adopting. When the share of the network adopting grows from near-zero to 10%, adoption propensity rises from about 20% to about 50%. This positive effect ceases beyond about 15% of the network, though we are not able to distinguish a flattening from a decline at high confidence.

This effect is consistent with our model's predictions in the following sense: if there are more adopters in an information-sharing network, then the support of an individual's observed $\boldsymbol{X}$ will tend to expand. This enables farmers to construct better models $\mathcal{F}(\boldsymbol{X})$ to explain the variance of $\Delta Y_{\mathbf{A}_i}$, thereby shrinking a farmer's estimate of the conditional variance and encouraging adoption. Furthermore, this may explain why the effect of the unconditional standard deviation is positive (if insignificant), because the number of adopters within the network is positively correlated with the variance of the outcomes – $\rho \approx .28$ in our case.

On the other hand, farmers who know one another but who do not talk about farming are responsive to the unconditional moments in their networks, and the marginal effects are somewhat stronger than they are in the full network. This is consistent with our conceptual model as well: when farmers do not know the $\boldsymbol{X}$'s that mediate the results that they see, their implicit estimate of the mean and variance of the outcome to adoption is equal to the unconditional moments of the outcome. Furthermore, the effect of the share of the network adopting is small and insignificant, which makes sense if our conceptual model is taken seriously; if the main avenue through which a larger number of observations affects the adoption decision is by increasing the support of $\boldsymbol{X}$ and thereby improving f, then more observations of y divorced from their corresponding $\boldsymbol{X}$ should have no effect on the expected utility of adoption.[15] Finally, the implied degree of risk aversion is substantially higher than in the estimates derived from the full network—the average marginal effect of a 1SD change in the standard deviation of observed outcomes is .39, about .17 percentage points larger than estimated from the full network.

Section A.5 explores several alternative specifications of the social network, including (*a*) sensitivity to the third moment of the observed outcome distribution, (*b*) a "good/bad news" specification inspired by Conley and Udry (2010), (*c*) a model in which farmers are sensitive to relative, rather than absolute changes, and (*d*) a model that excludes those who adopted before the 2014 short rains. Results are largely consistent across specifications.

Discussion

The RCT underlying this article originally sought to test biochar's profitability in smallholder agriculture, and to test means of stimulating its dissemination conditional on that profitability. Ultimately, biochar's profitability at sustainable prices was marginal, and uptake was low even with steep subsidies. Given the low adoption at prices consistent with cost recovery—let alone financially sustainable provision by a profit-making business—biochar may not be viable here absent a model that can cut the costs of its production and distribution. While biochar provides public goods such as carbon sequestration, it is not clear that internalizing this value would subsidize biochar enough to promote widescale uptake.

Biochar's failure to take hold, even under high subsidies, is viewed here as a specific case of a general problem—technologies that appear beneficial often do not spread virally. This article adds important texture to the broader literature on this phenomenon, which has found that heterogeneity impedes social learning in developing-world

[15] Except perhaps through the variance of the expectation (Foster and Rosenzweig 1995), though we do not find evidence for that here. We note that these results show the proportion of the network adopting biochar. Similar results are obtained using the total number of network contacts adopting, but that term is endogenous given that different farmers will have different-sized networks. We thus use proportions in order to avoid controlling for network size, which is nearly collinear with the number in the network adopting.

agriculture (Munshi 2004; Conley and Udry 2010; Tjernström 2015). Agricultural technologies are not uniformly profitable (Suri 2011), but rather their profitability is mediated by economic (Conley and Udry 2010) and environmental (Marenya and Barrett 2009a, 2009b) factors.

Given this heterogeneity, it is natural that learning about a technology's benefit is characterized by a process of constraining the space of potential outcomes. If prospective users do not immediately know their own potential outcomes, they are likely to infer their benefits from the experiences of their neighbours. If they do not know the mediators of the outcomes they observe, then the implicit conditional variance of the prospective outcome will be large, and this may inhibit adoption.

We find support for the proposition that social learning can be characterized by implicit model-building, rather than herding behaviour around unconditional outcomes. Rather than simply aping fellow farmers who obtain good results from technology adoption, farmers are rather more sophisticated. They behave like statisticians—observing factors that move together with outcomes, and using these data to build a model to explain and predict outcomes in new settings of their own. We also find, like others (Duflo, Kremer, and Robinson 2011a), that information flows about agricultural technologies are imperfect—not all farmers routinely or deeply discuss farming. Finding ways to stimulate information exchange could thus stimulate technology dissemination generally, while ensuring that adopting farmers are those who most stand to benefit.

Future work could build on these findings. To what degree could dissemination be accelerated by priming farmers to observe factors that are thought to be important (Hanna, Mullainathan, and Schwartzstein 2014)? If farmers are sophisticated model-builders, then to what degree is slow dissemination of agricultural technology in sub-Saharan Africa the result of technologies whose costs simply do not outweigh their benefits, given farmer's preferences and contexts? Future work could also investigate the degree to which social networks themselves change in response to technological change, and how this process unfolds to either inhibit or speed technology diffusion (Acemoglu, Bimpikis, and Ozdaglar 2014).

Supplementary Material

Supplementary material are available at *American Journal of Agricultural Economics* online.

References

Acemoglu, D., K. Bimpikis, and A. Ozdaglar. 2014. Dynamics of Information Exchange in Endogenous Social Networks. *Theoretical Economics* 9 (1): 41–97.

Bandiera, O., and I. Rasul. 2006. Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal* 116 (514): 869–902.

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson. 2013. The Diffusion of Microfinance. *Science* 341 (6144): 1236498.

Blundell, R., and J.L. Powell. 2003. Endogeneity in Nonparametric and Semiparametric Regression Models. *Econometric Society Monographs* 36: 312–57.

Blundell, R.W., and J.L. Powell. 2004. Endogeneity in Semiparametric Binary Response Models. *The Review of Economic Studies* 71 (3): 655–79.

Breiman, L. 2001. Random Forests. *Machine Learning* 45 (1): 5–32.

Brown, R. 2009. Biochar Production Technology. In *Biochar for Environmental Management: Science and Technology*, 127–46. London: Earthscan.

Burke, M. 2014. Selling Low and Buying High: An Arbitrage Puzzle in Kenyan Villages. Technical Report.

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

Chandrasekhar, A., and R. Lewis. 2011. Econometrics of Sampled Networks. Technical Report. MIT Working Paper.

Conley, T., and C. Udry. 2010. Learning about a New Technology: Pineapple in Ghana. *The American Economic Review* 100 (1): 35–69.

Crane-Droesch, A., S. Abiven, S. Jeffery, and M.S. Torn. 2013. Heterogeneous Global Crop Yield Response to Biochar: a Meta-Regression Analysis. *Environmental Research Letters* 8 (4): 044–049.

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM-Algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*: 1–38.

Dercon, S., and L. Christiaensen. 2011. Consumption Risk, Technology Adoption and Poverty Traps: Evidence from Ethiopia. *Journal of Development Economics* 96 (2): 159–73.

Dong, Y., and A. Lewbel. 2015. A Simple Estimator for Binary Choice Models with Endogenous Regressors. *Econometric Reviews* 34 (1-2): 82–105.

Duflo, E., M. Kremer, and J. Robinson. 2008. How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya. *The American Economic Review* 98 (2): 482–88.

———. 2011a. Learning by Doing but not Learning from Others: Experimental Evidence from Maize Farmers in Kenya. Technical report, Working Paper, Washington DC.

———. 2011b. Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *The American Economic Review* 101 (6): 2350–90.

Foster, A., and M. Rosenzweig. 1995. Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy* 103 (6): 1176–209.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. *The Elements of Statistical Learning*, Volume 1. Springer Series in Statistics Springer. Berlin.

Hammes, K., M. Torn, A. Lapenas, and M. Schmidt. 2008. Centennial Black Carbon Turnover Observed in a Russian Steppe Soil. *Biogeosciences* 5: 1339–50.

Hanna, R., S. Mullainathan, and J. Schwartzstein. 2014. Learning Through Noticing: Theory and Evidence from a Field Experiment. *The Quarterly Journal of Economics* 129 (3): 1311–53.

Hastie, T., and R. Tibshirani. 1990. Generalized additive models. John Wiley & Sons, Inc.

Heckman, J.J. 1978. Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica* 46 (4): 931–59.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry. 2014. Agricultural Decisions after Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics* 129 (2): 597–652.

Keiluweit, M., P. Nico, M. Johnson, and M. Kleber. 2010. Dynamic Molecular Structure of Plant Biomass-Derived Black Carbon (Biochar). *Environmental Science & Technology* 44 (4): 1247–53.

Kremer, M., and E. Miguel. 2007. The Illusion of Sustainability. *Quarterly Journal of Economics* 122: 1007–65.

Kuzyakov, Y., I. Bogomolova, and B. Glaser. 2014. Biochar Stability in Soil: Decomposition During Eight Years and Transformation as Assessed by Compound-Specific 14c Analysis. *Soil Biology and Biochemistry* 70: 229–36.

Lehmann J. 2007. A Handful of Carbon. *Nature* 447 (7141): 143–44.

Lehmann J., and S. Joseph. 2009. *Biochar for Environmental Management: Science and Technology*. Earthscan, London & Sterling, VA. 416p.

Lewbel, A. 2000. Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables. *Journal of Econometrics* 97 (1): 145–77.

Lewbel, A., Y. Dong, and T.T. Yang. 2012a. Comparing Features of Convenient Estimators for Binary Choice Models with Endogenous Regressors. *Canadian Journal of Economics/Revue Canadienne D'économique* 45 (3): 809–29.

Lewbel, A. 2012b. An Overview of the Special Regressor Method. Technical Report.

Li, Q., and J.S. Racine. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Lin, W., and J.M. Wooldridge. 2015. On Different Approaches to Obtaining Partial Effects in Binary Response Models with Endogenous Regressors. *Economics Letters*. 134: 58–61.

Manski, C.F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* 60 (3): 531–42.

Marenya, P., and C. Barrett. 2009a. State-Conditional Fertilizer Yield Response on Western Kenyan Farms. *American Journal of Agricultural Economics* 91 (4): 991–1006.

———. 2009b. Soil Quality and Fertilizer Use Rates among Smallholder Farmers in Western Kenya. *Agricultural Economics* 40 (5): 561–72.

Marra, G., and R. Radice. 2001. A Flexible Instrumental Variable Approach. *Statistical Modelling* 11 (6): 581–603.

Mullainathan, S., and J. Spiess. 2017. Machine Learning: an Applied Econometric Approach. *Journal of Economic Perspectives* 31 (2): 87–106.

Munshi, K. 2004. Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution. *Journal of Development Economics* 73 (1): 185–213.

Rubin, D.B., and N. Schenker. 1986. Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse. *Journal of the American Statistical Association* 81 (394): 366–74.

Ruppert, D., M. Wand, and R. Carroll. 2003. *Semiparametric Regression*, Volume 12. Cambridge University Press.

Saha, A. 1997. Risk Preference Estimation in the Nonlinear Mean Standard Deviation Approach. *Economic Inquiry* 35 (4): 770–82.

Sanchez, P.A. 1977. Properties and Management of Soils in the Tropics. *Soil Science* 124 (3): p.187.

Sanchez, P. 2002. Soil Fertility and Hunger in Africa. *Science* 295 (55–62): 2019.

Singh, N., S. Abiven, M. Torn, and M. Schmidt. 2012. Fire-Derived Organic Carbon in Soil Turns over on a Centennial Scale. *Biogeosciences* 9: 2847–57.

Sombroek, W.G., H. Braun, B. Van der Pouw. 1982. *Exploratory Soil Map and Agro-Climatic Zone Map of Kenya, 1980. Scale 1: 1,000,000.* Kenya Soil Survey.

Suri, T. 2011. Selection and Comparative Advantage in Technology Adoption. *Econometrica* 79 (1): 159–209.

Tjernström, E. 2015. Unpublished. *Signals, Similarity and seeds: Social Learning in the Presence of Imperfect Information and Heterogeneity.* Davis: University of California.

Vanlauwe, B., and K.E. Giller. 2006. Popular Myths around Soil Fertility Management in Sub-Saharan Africa. *Agriculture, Ecosystems & Environment* 116 (1): 34–46.

Wood, S. 2003. Thin-Plate Regression Splines. *Journal of the Royal Statistical Society (B)* 65 (1): 95–114.

———. 2006a. *Generalized Additive Models: An Introduction with R.* London and New York: Chapman and Hall/CRC.

———. 2006b. Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics* 62 (4): 1025–36.

———. 2011. Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society (B)* 73 (1): 3–36.

———. 2013. On P-Values for Smooth Components of an Extended Generalized Additive Model. *Biometrika* 100 (1): 221–8.

World Bank. World Bank Data. 2014. Available at: http://data.worldbank.org/.

Zimmerman, A. 2010. Abiotic and Microbial Oxidation of Laboratory-Produced Black Carbon (Biochar). *Environmental Science & Technology* 44 (4): 1295–301.