2021

Imperfect Information and Learning: Evidence from Cotton Cultivation in Pakistan

Amal Ahmad

Department of Economics, University of Massachusetts Amherst

Imperfect information and learning: Evidence from cotton cultivation in Pakistan

Amal Ahmad*

February 2021

Abstract

Information problems are pervasive in developing economies and can hinder productivity growth. This paper studies how much rural producers in developing countries can learn from their own experience to redress important information gaps. It builds a model of learning from experience and applies it using a rich dataset on cotton farmers in Pakistan. I test whether farmers learn from cultivation experience about the pest resistance of their seeds and use this information to improve selection and productivity. I find no such learning effect and this conclusion is robust to several parameters that could signal learning. The findings document the difficulty of parsing out and processing information from cultivation experience alone and point to the importance of information provision to producers by the government or external agencies.

1 Introduction

Economic development is a process characterized by potentially severe and persistent information failures for both private and public agents. Given the salience

*Department of Economics, University of Massachusetts Amherst. Email: amalahmad@umass.edu. I am grateful to Daniele Girardi, James Heintz, Peter Skott, Emily Wang, Ina Ganguli, Arslan Razmi, Mariam Majd, and two anonymous referees for very helpful comments on earlier drafts of this paper, as well as the participants of the Analytical Political Economy Workshop. I also wish to thank David Spielman of the International Food Policy Research Institute for graciously answering questions about the dataset. All remaining errors are my own.

of agricultural production in developing countries, the information failures faced by farmers are particularly important to understand. A large literature exists on the information problems that farmers in Africa, South Asia, and other parts of the world face in securing credit (Ghosh et al, 2000), in managing risk (Poole, 2017), and in learning and adapting agricultural technologies (Foster and Rosenzweig, 2010).

This paper contributes to the latter literature, on how producers in developing countries learn to adapt and effectively use technology, with a focus on imperfectly known seed-based technologies. Seed technologies arise out of mechanical hybridization or lab-based genetic engineering and can improve characteristics such as resistance to pests, reduction of spoilage, or nutrient profile. Developing countries account for the majority of GMO crop production in the world in terms of acreage and production (ISAAA, 2017) but regulatory mechanisms in these countries are notoriously weak including around seed assurance and quality control standards (FAO, 2009). Combined with the inherent information problem that one cannot deduce the attributes of a seed by physical inspection, and the compounded problem that much of these technologies originate from non-local expertise, this can create significant difficulties for farmers in selecting and cultivating high-yield crops.

I investigate whether, amid imperfect information, farmers can discover the "hidden" attributes of their seed from cultivation outcomes, since learning from experience is particularly valuable when sources of external information are limited.

I first provide a simple theoretical model in which an agent can learn about a profit-maximizing attribute from cultivation experience and uses this to enhance variety selection in the next period. The model elaborates this strategy and demonstrates the conditions under which the quality of the crop on the market improves and monetary benefit to farmers are generated.

After modelling the behavior that results from learning, I use it to derive a specification to test empirically for learning and apply it to a rich panel dataset on cotton cultivation in Pakistan. The difficulty in testing for learning from own experience is that the information must be inaccessible to the farmer somehow, so that there is space for learning and discovery, but accessible to the researcher to allow them to verify whether the right information was learned. This is the opportunity provided by the unique structure of the dataset I use, the Pakistan Cotton Survey (PCS).

Pakistan's cotton farming industry is an apt context to study learning about unknown seed characteristics because farmers have limited information about an important pest resistance technology of the seed varieties they purchase. Particularly, while farmers are well aware that some cotton varieties may have "Bacillus thuringiensis" (Bt), a gene biotechnology that emits toxins lethal to bollworm pests, they do not know at purchase point *which* varieties or packages have Bt. This information problem is due to issues in technology adoption upstream and to poor labeling of packages.[1]

Using a representative sample, the PCS survey team tested the level of the Bt protein in individual farmers' plots in 2013 and only revealed the results to them two years later, enabling me to use farmer behavior and decisions between 2013 and 2014 to study whether farmers learned from cultivation about information that was unavailable to them ex-ante. I can test whether farmers, after observing the performance of their 2013 crop, accurately assess the pest resistance of the variety and whether they respond as predicted by the model, switching varieties next year if they learned that their variety was lacking in Bt pest resistance and vice versa.

[1] As explained in Section 2, farmers adopted the Bt gene, which was engineered by Monsanto in the US in 1996, haphazardly, through unlicensed borrowing of the original variety and mixing with local varieties (Speilman, 2017). It was this haphazard adoption process, coupled with the weak capacities of the Pakistani state in tracking and labeling varieties, that generated the information problem in the market.

The results show that farmers are unable to learn about biophysical resistance by observing cultivation outcome. Specifically, a key finding is that the actual pest resistance of the seed variety employed in season t by a farmer does not predict the probability of seeking a different seed variety in $t + 1$. Additional results show that lack of learning behavior arises because it is difficult for farmers to distinguish whether poor pest resistance performance is driven by the biophysical characteristics of the plant (low Bt) or by unfavorable environmental conditions. Specifically, I find that the probability of switching seed variety is significantly influenced by the farmer's perception of pest resistance, with farmers who assess (post-harvest) resistance as lower being more likely to switch next period, but that these perceptions of pest resistance are not correlated with actual Bt resistance. In addition to these results, I check that farmers do not learn about Bt content but react in ways other than variety switching, such as increased pesticide use with low Bt varieties.

Since cultivation experience is not sufficient to redress the information gap, the results suggest that policy, in the form of stronger certification standards by the government or information provided externally to farmers by agricultural extension services, might be necessary for farmers to make more informed choices.

A rough back-of-envelope exercise suggests that the lack of learning I document in this paper leads to large productivity losses. Based on the size of Pakistan's cotton cultivation industry and the documented effects of Bt on damage abatement, I estimate that failing to learn about Bt content and to purchase maximum effectiveness seeds results in long term losses up to 170 million USD, or 12.5% of industry value in 2013-2014. Therefore, this paper has implications not only for microeconomic behavior but also for productivity growth at the industry level.

This study contributes to three related but distinct strands of literature. First it sheds light, theoretically and empirically, on how producers may use own experience to learn under imperfect information. The literature on agricultural producers in developing countries has more commonly explored learning from external information, typically from extension services (Murphy, 2017; Emerick et al, 2016; Maertens et al, 2018), or from social networks (Munshi, 2004; Conley and Udry, 2010; Crane-Droesch, 2017). This paper instead focuses on the ability of farmers to uncover information organically, without the aid of externally verified information and through own experience. Own experience is important to understand because external information provision is rare and often expensive,[2] and because heterogeneity in growing conditions can mute social learning or peer effects (Foster and Rosenzweig, 2010).

Within the literature on learning from own experience in rural parts of the developing world, this paper complements the findings of, but is distinct from, Hanna et al (2014) and Bold et al (2017). In Hanna et al (2014), Indonesian seaweed farmers deal with a traditional technology, pod size, on which information can be readily available but which they fail to notice because they do not know the significance of pod size for yield and because there are many competing demands on their attention. By contrast, this paper deals with a relatively new biotechnology, whose significance the farmers (from their survey answers) clearly comprehend but whose facets can be very difficult to deduce from production experience despite exerting the effort to notice. In Bold et al (2017), Ugandan maize farmers deal with unknown levels of fertilizer effectiveness, which they have trouble learning about due to noisy yield signals; the findings are generated by calibrating a learning model to outcomes from researcher-managed experimental plots to simulate what farmers would or would not learn. The findings of my paper also suggest that noisy yield signals can make learning from cultivation

[2]The lack of focus on extension services is particularly appropriate for this study; the farmers indicate the near absence of any help from NGOs, farmer cooperatives, or other extension services.

experience very difficult, but I test for learning by applying theory directly to farmer behavior in the field.

Second, this study provides insights on consumer learning when goods' attributes are hidden or not easily observed. It demonstrates whether key attributes of an important commodity, agricultural seeds, can be evaluated by the consumer (farmer) after experience/use or if these attributes cannot be revealed even after use. The literature on this subject terms the former an experience good and the latter a credence good (Darby and Karni, 1973; Girard and Dion, 2010). In this paper, genetically modified seeds that are not properly labeled are either experience goods, if farmers can learn about their attributes from experience, or credence goods, if they cannot evaluate said attributes even post-experience.

Therefore, the main question in this paper can be reformulated as an inquiry into the information-characteristics of a key commodity in rural developing markets.[3] Since the government can greatly ameliorate the information problem for consumers if it provides credible labeling and certification (Dulleck et al, 2006; Dulleck et al, 2011), the paper also demonstrates the consequences of weak government capacities and high costs of certification for developing-country agents facing information problems.

Third, the paper contributes to the development literature more broadly by demonstrating how information problems generated in the technology acquisition stage in a development context can trickle down and hinder effective use after adoption. The information problem in this case emerged during the acquisition of the Bt gene, due to constraints on effective local adaptation and governance.[4] Challenges with tech-

[3]Few studies address the credence goods problem in developing countries; none except Auriol and Schilizzi (2015) focus on agricultural seeds. Even that paper is a theoretical investigation of the costs of certification, not an empirical application.

[4]Agents in developing countries are also innovative and constraints do not imply lack of agency. The local mixing of the Monsanto protein with the local germplasm, while haphazard, afforded

nology import and local adoption are widely acknowledged in development economics (Dosi, 1988; Bardhan and Udry, 1999; Khan, 2010) but it is unclear how much information failures generated at that stage persist post-acquisition. This paper's results demonstrate high persistence in one such market.

In turn, high persistence can point to potential spillage into other markets and the deepening of other information problems. For example, in rural financial markets, the agent, if borrowing to purchase inputs, may face difficulty evaluating input quality and the ability to pay back the loan. In this case, incentive-compatible mechanisms to overcome principal-agent problems will not be sufficient to give the lender all the relevant information. Missing information in developing countries is often not strategically hidden but unknown, and corrective strategies must operate accordingly.

The paper is organized as follows. Section 2 provides background to the information problem in the Pakistani cotton seed market. Section 3 builds a model of learning from experience and response behavior, and shows the relationship between information costs, learning, and market outcomes. Section 4 describes the dataset. Section 5 outlines the econometric methodology derived from the theoretical model and explains the identification strategy and sample selection. Section 6 presents and discusses the empirical results. Section 7 considers and rules out alternative explanations of the findings and offers robustness checks. Section 8 summarizes and concludes.

---

the farmers stronger pest resistance into their crop that they would not have had otherwise. The Pakistani state, though it struggled with regulating the seed market, used its power to prevent Monsanto from pushing for a patent in Pakistan, affording farmers the space to create local hybrids under legal cover.

2 Background

Producing around 8 million 500 pound bales per year, Pakistan is the fourth largest producer of cotton in the world and also its fourth largest exporter after China, the US, and India. In 2019, it was estimated that over 1.6 million farmers cultivate cotton in Pakistan, with cotton cultivation accounting for 15% of all arable land during the Kharif (April-July) season and 26% of all farms in the country. The downstream textile industry is also integral to the country's economy, employing about 10 million people and generating 50% of all foreign exchange (USDA, 2019).

Pakistan's cotton farmers, based almost completely in the Punjab (75%) or Sindh (24%) provinces, have increasingly adopted the genetically modified bollworm-resistant[5] Bacillus thuringiensis (Bt) cotton over the past fifteen years, and evidence suggests that Bt use has reduced crop damage and improved yield (Ali and Abdulai, 2010; Kouser and Qaim, 2013). However, the way in which Bt has been adopted has been haphazard and largely unregulated. Bt cotton can rely on different cry proteins to generate toxins that confer the bollworm-resistance criterion, but the majority of Bt cotton varieties in Pakistan "rely on the cry1Ac gene from the MON-531 event developed by Monsanto [in 1996]." (Spielman et al, 2017; p.2) In the mid-2000s, lacking a formal system for proper Bt-variety acquisition due to Monsanto's iron-clad patents,[6] Pakistani farmers began introgressing this specific gene into local germplasm to create locally specific hybrid Bt varieties. Local Pakistani farmers were hence able to use trial and error and mixing with local germplasm to "effectively" introduce Bt to their cotton crop, despite intellectual property barriers.

[5]A bollworm is a moth larva that attacks cotton and is a major pest concern for producers.

[6]Monsanto had patents in the US but not Pakistan; it tried very hard to obtain a patent in Pakistan after realizing local farmers were introgressing the cry1Ac gene but the Pakistani government refused to grant it one.

Since adoption, the release and marketing of Bt cotton has been largely unregulated in Pakistan. Seed varieties are often missing labels or contain incomplete or unregulated labelling. There is a lack of "regulatory systems.. [to properly] enforce rules requiring seed sellers to provide technical information on quality alongside their product.. [and] the judicial system does not provide sufficient recourse for farmers defrauded by seed sellers" (Spielman et al, 2015; p.1). Due to the inherent information problem in seed markets (a farmer cannot look at a seed and infer its quality), farmers are subject to a serious information asymmetry when purchasing seeds in the absence of proper regulatory mechanisms.

Local mixing, which can result in poor breeding methods or improper genetic checks, and poor regulatory capacities have resulted in the promulgation of low-quality seed-based technologies in Pakistan's cotton seed market. In a survey of 20 districts in 2008-2009 with farmers who thought they were planting Bt cotton, Ali et al (2010) found that 10% of the samples from Punjab did not test positive for the cry1Ac gene and of those that tested positive, only 36% contained concentrations sufficiently lethal to kill bollworms; the numbers were 19% and 41% for samples from Sindh. In a later study on the 2011 season, Ali et al (2012) used different technology on another sample and found that 30% of all varieties tested were not positive for any cry gene.[7]

The survey team that gathered the dataset on which this paper draws, the Pakistan Cotton Survey 2013-2014, sheds more light on these issues through two main papers. In Spielman et al (2017), the authors compare what the farmers are really planting to what they think they are planting. They find that a large portion of

[7]These results echo earlier findings about China, with Pemsl (2005) highlighting the lack of regulation, ubiquity of information imperfections, and subpar Bt effectiveness in China's Bt cotton seed market at the time.

farmers particularly in Punjab believe they are planting Bt when their variety is not actually Bt effective. They also run a logit model to predict the inaccuracy of belief and find the only significant predictor is education, with more educated farmer less likely to hold erroneous beliefs. However, they do not test for learning by linking Bt content with possible behavioral outcomes in the next season that could signal learning, as this paper does. In Ma et al (2017) the authors explore the cotton yield of the sampled farmers and find that, in a nonlinear damage abatement model, Bt effectiveness as measured by the PCS has a significant positive effect on farmer yield, when other input use is controlled for.

3 Theoretical model

We expect farmers who learn to behave after discovery in ways that reflect their knowledge. With seed-based technologies, one possibility is that farmers alter the variety they purchase next year, with those who discover their variety was high in that attribute being more likely to repurchase it, other factors constant, and vice versa. I illustrate this response strategy and how it can be affected by the costs of gathering and processing the relevant information. I also show the conditions under which learning improves market outcomes, in terms of the average attribute level on the market. In Section 5, I use this theoretical model to derive econometric specifications to test for learning from experience.

Suppose an observable outcome for farmer i at time t, Y_{it}, is a function of the unknown level of some attribute B_{it} and of other factors e_{it}, so that $Y_{it} = f(B_{it}, e_{it})$. In this case, for example, Y_{it} could be pest damage. Farmers may discover B_{it} ex-post (in $t+1$) if f is known and e_{it} is easily observable, so that B_{it} is deduced by exclusion. Conversely, if it is difficult to know f or observe e_{it} or both, then discovering B_{it} ex-post is less likely. This deduction is not necessarily a costless process, as I discuss below.

Let there be two periods $t = 1, 2$ and let B_t denote the price-adjusted level of a profit-enhancing attribute in period t. In this case B_t is the Bt level in the seed variety per rupee spent on the variety, but the model can apply more generally to other markets and attributes. For Pakistani cotton farmers, Bt content as (one) driver of variety selection is plausible since the farmers cite bollworm-toxicity as important in their seed selection process. It should be noted that the farmers (from their answers) do not store cotton seed for use in the next cultivation period; those who report cultivating the same variety in 2014 bought that variety again in 2014.

I assume there is a market surplus each period, with more seeds available for sale than being bought. Specifically, there is general excess supply, so that a farmer can select any variety in either period. Though somewhat stringent, this assumption is backed by the responses of farmers in the survey, who suggest there is easy access to seeds and that seed prices are not at all prohibitive.

Excess supply also suggests that demand shifts in the second period can be met without a large relative change in prices, so that high-yielding varieties do not become too expensive and hence less desirable. Even if the relative price of in-demand varieties increases, as long as the relative Bt differential is still higher, the qualitative conclusions of the model hold. To simplify, I assume that the relative prices of different varieties are fixed between the two periods.[8]

Let the Bt of seeds for sale in the first period B_1 be a random variable distributed

[8] If we relax the assumption of excess supply so that some of the high Bt seeds become less available next period, the qualitative results of the model hold but the extent of switching and benefits from learning decreases. The "real-world" would lie somewhere between the scenario of too-low supply where farmers cannot purchase a better variety if they wanted to, and this opposite limiting case of general excess supply.

normally at $(E(B_1), \sigma^2)$. To differentiate between varieties consumed and the wider supply pool, I will notate the Bt level of varieties consumed with a tilde, as $\tilde{B}_t$. Due to the pervasive information problem, when farmer i purchases a variety in the first period, they receive a single random realization $\tilde{B}_{1i}$. They cannot identify $\tilde{B}_{1i}$ at purchase point due to poor labeling and certification standards; while on average the farmer receives the mean level on sale, so that $\tilde{B}_{1i} = E(B_1)$, what each farmer actually gets deviates from this amount by a random error component and may be above or below the market average.

However, while the farmer does not know $\tilde{B}_{1i}$ (what they are getting), they have an expectation, V_1^*, about it at purchase point. I assume all farmers who think they are purchasing Bt share the same ex-ante expectation (I address the importance of fixed expectations in the empirical section). It is possible that expectations correspond to the mean quality in supply, so that $V_1^* = E(B_1)$, or that there is systemic error in the farmer's assessment, $V_1^* = (B_1) + \gamma$. In the second period, seeds available for sale have Bt level B_2 which is a random variable with the same distribution as the year prior, $E(B_2) = E(B_1)$.[9].

Given the persistent absence of certification standards, producers that switch varieties from t_1 to t_2 will simply be going back to the supply pool and picking at random from it once more. Letting s be the switching decision, then:

$$E_i(\tilde{B_{2i}}) = E(B_2) \quad [= E(B_1)] \qquad \text{if} \qquad s_i = 1 \tag{3.1}$$

For those who do not switch varieties, in a perfect market, buying the same variety again would mean getting exactly the same Bt content again: $\tilde{B_{2i}} = \tilde{B_{1i}}$, so that at

[9]If producers do not offer the varieties that "did not sell" in the previous season, so that $E(B_2) = E(\tilde{B}_{1i})$, this still holds since $E(\tilde{B}_{1i}) = E(B_1)$

least those who"stick" with their old varieties would no longer have an information problem once they "discover" the Bt content of one package. However, given that the varieties are poorly labeled and certified, it is possible that something being sold as the same variety actually has a different level of Bt in the next season. Let p be the probability that the farmer gets the same Bt content again if they do not switch (variety integrity), and $1-p$ be the probability that they get something completely random from the overall pool even though the variety is being marketed as the same one.[10] Then for those who do not switch, their expected second-period Bt content will be

$$\begin{aligned} E_i(\tilde{B_{2i}}) &= p\tilde{B_{1i}} + (1-p)(E(B_2)) \\ &= p\tilde{B_{1i}} + (1-p)(E(B_1)) \qquad \text{if} \quad s_i = 0 \end{aligned} \tag{3.2}$$

To see when producers switch, we note that profit is a positive function of price-adjusted pest resistance: $\pi_t = \pi(B_t)$, where $\pi' > 0$. Farmers will only switch varieties if they believe expected content next period with switching, V_1^*, is greater than expected content without switching, $p\tilde{B_{1i}} + (1-p)(V_1^*)$. So, $s = 1$ only if $V_1^* > p\tilde{B_{1i}} + (1-p)(V_1^*)$, or $V_1^* > B_{1i}$:

$$s_i = \begin{cases} 0 & \text{if } \tilde{B_{1i}} \geq V_1^* \\ 1 & \text{if } \tilde{B_{1i}} < V_1^* \end{cases} \tag{3.3}$$

Therefore, if farmers are able to discover Bt content from experience, they will switch varieties next year if Bt content this year fell below expectations and keep the same variety otherwise. This is represented in **Figure 1**:

[Figure 1 here]

[10]The higher p is, the more functioning the market is - variety names are meaningful. In the other extreme, if $p = 0$, a packet's variety name does not reflect a standardized variety at all.

This setup assumes that farmers can accurately pay attention to, identify, and act on the difference $\tilde{B_{1i}} - V_1^*$ even when that difference is very small. However, the is strong evidence that people do not always use or act on available information because cognitive limitations make it costly to pay attention to, and process, information (Sims, 2003). This phenomenon of "rational inattention" suggests that optimizing agents may rationally ignore or not pay attention to information if the benefits are small relative to the cost of acquiring and processing it, especially when there are many competing demands on their attention. More concretely, as attention costs become very large, agents pick deterministically from an option that was best ex-ante; they do not appear to be optimizing even though they are acting rationally by taking cognitive costs into account. Only as attention costs go to zero do they pick the best option in that state, acting as would be expected by classical theory (Dean, 2019).

Farmers have numerous competing demands on their attention and need to make many decisions. Moreover, the relative cost versus benefit of exerting attention and processing information to uncover $\tilde{B_{1i}}$ and act accordingly may depend on the absolute difference $|\tilde{B_{1i}}$ - $V_1^*|$. It is likely that, as $\tilde{B_{1i}}$ is "extreme" (very high or very low), it is more immediately obvious or easier to parse out from other factors that affect pest damage. It can also be verified (below) that the benefit from subsequent switching increases as $\tilde{B_{1i}}$ is more extreme relative to V_1^*. Therefore, greater $|\tilde{B_{1i}} - V_1^*|$ would be accompanied by lower costs and higher benefits of gathering and processing the relevant information, and vice versa.

This suggests that as $|\tilde{B_{1i}} - V_1^*|$ falls, farmers are less likely to exert the sufficient (costly) effort to uncover $\tilde{B_{1i}}$ and more likely to simply choose an ex-ante best strategy, which is a tossup between switching or not. Conversely, as $|\tilde{B_{1i}} - V_1^*|$ increases, farmers are more likely to deduce $\tilde{B_{1i}}$ and act according to Equation (3.3). The result is that, if learning is possible, farmers are more likely to switch when Bt content is *much lower* than expected and more likely to keep the same variety when content is *much higher* than expected. Switching becomes probabilistic instead of discrete, and involves the smoothing of the curve in **Figure 1**, as shown in **Figure 2**. This smooth curve can have a general function for the probability of switching $Prob(S)$, so that $Prob(S) = g(\tilde{B_{1i}} - V_1^*)$ where $g' < 0$.

[Figure 2 here]

The rational inattention framework can also help explain the absence of learning. If learning is impossible, the idea is that there are prohibitive cognitive limitations on the economic agent - that nobody can observe Y_{it} and deduce $\tilde{B_{1i}}$, perhaps because the effects of other confounding environmental factors are hard to separate out (i.e. e_{it} is impossible to observe or measure). In this case the attention costs needed to parse out the relevant information are infinitely large and farmers are unable, at all points, to discern Bt content and to act accordingly. The slope would be flat and farmers are most likely to choose the ex-ante best strategy (tossup) at each realization, so $g' = 0$, as shown in **Figure 3**.[11]

[Figure 3 here]

This model is useful not only for conceptualizing the learning process, but also for estimating the benefits to industry from such a process. In **Appendix A**, I show that if g takes a simple linear form, then it is easy to calculate how much learning (or lack thereof) about Bt content helps (or harms) industry revenue.

[11]This is distinct from the failure to notice that results from misunderstanding the attribute or its relevance to production, as in Hanna et al (2014). The Pakistani cotton farmers surveyed in PCS understand what Bt is, know how long they have been purchasing (what they think are) Bt varieties, and can identify their varieties' complex titles.

Specifically, I assume a linear form $P(S) = -\alpha(\tilde{B}_{1i} - V_1^*) + g_0$, where $\alpha > 0$;[12] the parameter α is the learning (and response) parameter. It measures the extent to which farmers act, through variety selection in the next period, on the deduced difference between Bt content and expectations. The intercept g_0 captures the rate of switching when values matches expectations; it can be 0.5 as in **Figures 2-3** to generate a tossup, or it can be any other constant capturing the effect of other variables on switching; this value does not affect any results. I also still allow for a discrepancy between expectations and true market averages, $V_1^* = E(B_1) + \gamma$.

When these functional forms are used to calculate the expected change in Bt content for each farmer with initial realization $\tilde{B}_{1i}$ from t_1 to t_2 based on their probabilistic switching decision, and to sum across all farmers to find the expected change in Bt content averaged across the market, it can be shown (detailed calculations in **Appendix A**) that:

$$E(\Delta\tilde{B}) = \alpha p \sigma^2 \tag{3.4}$$

The model therefore shows that as α, the extent of learning and response, and p, the extent of variety integrity, increase, average Bt quality consumed rises in the second period; this improvement is greater the larger the variance of Bt in the population. Conversely, if there is no learning, $\alpha = 0$ (or no variety integrity, $p = 0$) then average pest resistance is stagnant. Importantly, these conclusions are not affected by farmer expectations: as long as the farmers have a uniform ex-ante expectation V_1^* it does not matter that this expectation is accurate on average ($\gamma = 0$) or not. The probability of switching may also depend on the standardized deviation from

[12]The function would be bounded between 0 and 1.

expected content:

$$g(x) = -\alpha \left[\frac{\tilde{B}_{1i} - V_1^*}{\sigma} \right] + 0.5 \tag{3.5}$$

In that case, Equation (3.4) is amended as follows:

$$E(\Delta \tilde{B}) = \alpha p \sigma \tag{3.6}$$

With the average change in consumed Bt in the market $E(\Delta \tilde{B})$ in hand, we can, with additional information on the effect of Bt on yield, estimate the extent of monetary benefit to farmers from said improvement in average Bt. **Appendix B** illustrates how different values of α and p translate to expected Bt improvement $E(\Delta \tilde{B})$ for a fixed σ, and how, using evidence-based benchmarks on the effect of Bt on yield and revenue, this would then translate into sizable revenue gains for Pakistani cotton farmers in one year.

In Section 5, I use Equation (3.5) to derive an empirical specification to test the value of α or the negative of the slope of the curve in **Figure 2**, linearly approximated. This would allow us to test for the presence of learning. With this estimate, and for given values of p and σ, we can also infer how much average Bt would improve from one period of learning and switching, and estimate monetary gains.

There is a qualifier to this approach. If tests show that α is positive (**Figure 2**), we can conclude that farmers learn from experience and respond accordingly. However, while absence of learning necessarily generates a zero slope (**Figure 3**), the converse is not always true: having a zero slope or null coefficient does not necessarily imply farmers have not learned. There remains the possibility that farmers are able to gauge Bt levels from observing the crop's pest resistance but do not respond with switching varieties, and this would occur if they believe $p = 0$.

To see why it is rational for farmers to not respond to learned information through seed selection if they believe $p = 0$, note that the expected content from switching (Eq. 3.1) and from not switching (Eq. 3.2) become equivalent, so the profit maximization exercise that drives switching is invalidated.

Therefore, the empirical section will also gauge whether farmers believe $p = 0$ or not. It is not possible from the data to test actual variety integrity, but it is possible to gauge whether farmers believe switching varieties in response to low resistance is an effective strategy i.e. if they believe that $p > 0$. Only then can we interpret a null α coefficient (flat slope) as absence of learning.

Finally, I can check the underlying logic of the model and learning more directly. Farmers learn when attention and information costs are low enough to discern the effects of the plant's biophysical properties on resistance performance from the effects of other factors. A positive α should be accompanied by analysis showing that higher Bt content improves perceptions of the variety's bollworm-resistance, while with a null α we would not expect this relationship.

4 Data

I use data from the Pakistan Cotton Survey, which consists of four sequential in-person surveys and one biophysical sample survey. The surveys were conducted by the International Food Policy Research Institute (IFPRI) along with local agricultural scientists between March 2013 and January 2015, on a random stratified sample of farmers in Punjab and Sindh. These provinces account for 99% of all cotton production in the country, and the sample is nationally representative.[13]

The first survey, Round 1.1, collected preliminary background data on 727 cotton

[13] The surveys are accessible publicly from the Harvard Dataverse website.

farmers through face to face interviews in March 2013, prior to the beginning of sowing for the year. The farmers were asked about their personal and farming background and history and various plot characteristics.

The second survey, Round 1.2, followed up with the farmers in October 2013 after seeds were sowed, and only 601 of the farmers ended up sowing cotton for the season, so this represents the actual "base point" for the possible sample. Farmers were asked, among other things, about the variety purchased, whether they think their variety is Bt, cotton cultivation by plot, input use (water, fertilizer, and pesticides), and access to social networks and to credit.[14]

The third survey, Round 1.3, followed up in January 2014 and at this time the last picking for the season (harvest) was complete. The farmers were asked about input use, quantities harvested and sold, revenue, and perceptions about the performance of the crop. They were also asked about assets owned, general consumption patterns, and decision-making by gender.[15]

The fourth survey, Round 2.1, went back to these farmers in January 2015 and asked farmers the same questions as in Rounds 1.1-1.3, but this time for the 2014 harvest. The number of participants narrows further, as only 501 of those who cultivated cotton in 2013 also did so in 2014.

The Biophysical Sample Survey took place in July and August of 2013, between Round 1.1 and Round 1.2. Unlike the above, which were in-person interviews lasting hours at a time, this survey involved the team first obtaining the farmer's consent and then, for those who sowed cotton in 2013, randomly selecting a few cotton leaves and bolls at 70 and 120 days after sowing. The samples were taken to national laboratories

[14] Farmer answers show that social networks such as farmer coops are nearly nonexistent and that the use of cash credit is negligible.

[15] Nearly all decision makers in the dataset are male.

where they were tested for the presence of specific genes and toxins that contribute to Bt expression; the methodology is detailed in Ma et al (2017).

Crucially to this study, the farmers were not made aware of the biophysical sample results for the 2013 crop until early 2015, by which point the 2014 growing season was also finished.

Figure 4 illustrates the timeline of the surveys and corresponding cultivation stages. To my knowledge, this dataset has not been utilized beyond the studies conducted by the survey teams in Spielman et al (2017) and Ma et al (2017).

[Figure 4 here]

5 Econometric methodology

5.1 Specifications

To measure the extent of learning α, I regress variety switching in the next year on standardized Bt content, or $[(\tilde{B}_{1i} - E(B))/\sigma]$, in the current year. This derives directly from the specification in Equation (3.5), and generates a regression coefficient that is the slope of the linearized function in **Figure 3.2**. Bt content is seed-price adjusted by including seed price as a control in the regression. It does not matter whether or not $E(B_1) = V_1^*$ since subtracting any constant from the numerator does not affect the value of the regression coefficient. By contrast, a heterogenous V_1^* would require farmer-specific fixed effects for empirical assessment, untenable in this dataset because this is not multi-year panel data.

To ensure the assumption of homogenous expectations V_1^* holds, I principally use observations on farmers who when they bought the seed said they believe it is a Bt seed (which is the majority of farmers). This would roughly fix for ex-ante expectations. However, I also check that the inclusion of observations from farmers who thought they were not purchasing Bt and from farmers who did not know, and controlling for these as two other beliefs using dummy variables, does not change the results.

The main regression is:

$$Change_i = \beta_0 + \beta_1 BtLevel_i + \sum \beta_j Controls_{ji} + \epsilon_i \tag{5.1}$$

Change takes a value of 1 if farmer i switched varieties in 2014, and 0 otherwise. *Bt level* is the (standardized) Bt effectiveness of the farmer's 2013 variety as measured by the Biophysical Sample Survey. It is measured in micrograms of the relevant protein per gram of leaf tissue; a higher level indicates more toxin, therefore higher effectiveness in targeting and eliminating bollworms.[16] Controls are other factors, occurring in 2013 or beforehand and including seed price, that can affect variety change in 2014.

I expect $\beta_1 < 0$ if learning is present, with farmers who discover low Bt content more likely to switch and vice versa; β_1 is equivalent to $-\alpha$ in the theoretical model. Given that farmers did not have external information about the Bt content of their variety, any learning about this attribute reflected in an impact on purchase decisions in the next season would have been uncovered from cultivation experience. Conversely, if there is no learning, Bt of the 2013 variety would not affect seed choice the following year and I would expect $\beta_1 = 0$.

[16] For each farmer/variety, the survey team randomly collected 2 leaf and 2 boll tissues from the main plot, at both 70 days after sowing and 120 days after sowing, and measured the toxin expression for each of these in-lab using the ELISA sandwich test. My variable is an average of the measurements 70 days after sowing for each variety. The data for 120 days after sowing is less complete and has more variation per observation, but the results do not change even when I include it in the analysis.

Next, to verify that farmers would resort to switching if they learned about Bt content, I look at farmer perceptions. As explained in Section 3, farmers will only switch from low-Bt seeds if they believe there is some variety integrity in the market. In Round 1.3, immediately after the 2013 harvest was complete, farmers were asked to evaluate the bollworm resistance of their crop as poor, moderate, or very good. If farmers believe switching is an effective strategy for improving pest resistance (if they believe $p > 0$), we expect them to switch varieties in 2014 if they felt their 2013 variety had poor resistance, all else constant. The relevant regression is:

$$Change_i = \gamma_0 + \gamma_1 ResistancePerception_i + \sum \gamma_j Controls_{ji} + \epsilon_i \tag{5.2}$$

If farmers believe $p > 0$ we would expect $\gamma_1 < 0$: farmers who evaluate bollworm resistance as lower are more likely to change seed variety next year; they do think switching is an effective strategy for improving seed effectiveness. This would support the behavior outlined in the theoretical model, so that a null β_1 in Equation (5.1) would signal the absence of learning as opposed to farmer unwillingness to switch.

The perceptions variable can also be used to sharpen the insight on the learning process. Farmers learn about Bt content if they can distinguish the extent to which pest resistance performance is driven by the biophysical attributes of the plant versus environmental and other factors. We can regress perceptions on Bt content and on those controls:

$$ResistancePerception_i = \theta_0 + \theta_1 BtLevel_i + \sum \theta_j Controls_{ji} + \epsilon_i \tag{5.3}$$

$\theta_1 > 0$ would be a direct indication of learning, since it implies higher Bt content improves the farmer's perception of bollworm resistance. Therefore, we expect $\theta_1 > 0$ in Equation (5.3) to be associated with $\beta_1 < 0$ in Equation (5.1). Conversely, if learning is difficult, Bt effectiveness remains unknown because it is difficult to discern the effect of the biophysical attribute of the plant on performance ($\theta_1 = 0$). This would imply no impact of Bt on variety choice ($\beta_1 = 0$). In both scenarios, however, I expect $\gamma_1 < 0$ from Equation (5.2), to signify that farmers believe there is some market integrity and act as the model predicts.

Finally, it is possible to diverge from the theoretical model in Section 3 and test whether farmers respond to low Bt content by increasing pesticide use during cultivation instead of changing variety. The specification is:

$$Pesticide_i = \phi_0 + \phi_1 BtLevel_i + \sum \phi_j Controls_{ji} + \epsilon_j \tag{5.4}$$

Pesticide measures pesticide use per acre in 2013, constructed by adding the quantities of various pesticides and dividing by acres of cotton cultivated.[17] Learning would imply $\phi_1 < 0$, since farmers realize that the plant itself is emitting toxins lethal to pests so that they can use less pesticide. With no learning, ϕ_1 is close to zero and insignificant.

Table 1 summarizes the possible coefficient combinations and interpretations.

[Table 1 here]

5.2 Controls

In Equation (5.1) and (5.2) I control for other factors that can affect variety selection:

- Farmer characteristics that may influence how the farmer deals with their crop (education, years of general farming experience, years of experience cultivating

[17]There is not enough information for most observations to construct an "effective" pesticide measure that weighs quantities by percent strength.

what they think is Bt cotton, land owned as proxy for wealth).

- Planting history for the specific 2013 variety.
- Price per unit of the seed variety purchased for the 2013 season.
- Price per unit of post-harvest cotton fetched by the 2013 variety.[18]
- Input intensity (irrigation, Nitrogen fertilizer, labor, and seeds sowed, all per acre of cotton cultivated).
- Dummies for geographical district, since the observations belong to 22 districts, each of which share ecological and cultural properties that very likely affect cultivation attitudes.

The controls are all measured in 2013 or beforehand, hence predetermined relative to the dependent variable *Change*.

In addition to the above controls, for Equation (5.3), exogenous pest intensity affects resistance performance and should also be controlled for but no reliable information on this is available. Since it appears that pest intensity is time- and space-dependent, I assume that controlling for time-of-sowing and geographical district can be roughly sufficient. For Equation (5.4), I also control for soil type (since it can impact pesticide absorption).

Appendix C details how these control variables are constructed. It also illustrates their distribution, as well as the distribution of the key dependent variables, in the data.

[18]Cotton selling price is distinct from yield performance and captures desirable qualities such as whiteness of the cotton and quality of the lint.

5.3 Identification

For Equation (5.1), identification straightforward. First, reverse causation is ruled out because Bt level is measured for the 2013 variety while the choice to change varieties is made in 2014. Even without the time lapse, it is not clear how farmer choice can affect a biological characteristic of the crop which is not known in any verifiable way to the farmer themself ex-ante.

Second, controlling for farmer characteristics and farming experience enables me to control for factors that could affect both Bt level in 2013 (if it is not completely random) and the switching choice in 2014. If there *are* any unobserved factors that make farmers who end up with higher quality seeds in one year also be more prone to information discovery and learning, then the resulting bias would pressure the coefficient of interest upward. Therefore, a null result from Equation (5.1) is particularly resilient againt omitted variable bias.

Similary, Equation (5.2) involves measuring a clear directional relationship: the effect of perceptions formed after the 2013 harvest was completed, on purchase choices made in the following season. Appropriate controls on farmer characteristics and experiences help to control for potential confounding factors that could affect both perception formation and selection choices.

Equation (5.3) also incorporates a key independent variable (Bt level) which precedes the dependent variable (post-harvest farmer perceptions), as well as controls on the farmer characteristics, experience, and sowing conditions that may be correlated with Bt levels and affect post harvest perceptions.

In Equation (5.4) pesticide choices are made during cultivation, during which we would also expect any learning to take place. Therefore, the specification is only a valid test for learning if farmers can learn about Bt content before cultivation is over, so that there is room for adjusting pesticide decisions in the same season. It is not clear if this is the case or if input use in that season is predetermined relative to information learned later in the season about Bt content. Therefore Equation (5.4) is not the focus of the discussion but used as a supplemental result.

5.4 Sample

As shown above, though the initial pre-cultivation sample of farmers was larger, only 501 farmers cultivated cotton in both 2013 and 2014. Furthermore, among those, a number of farmers did not have Bt samples taken from their plots, or did not farm on the main plot on which sufficient information is available, or did not answer basic questions including on their farming experience. This narrows the number of farmers who cultivated cotton in both seasons, and on whom sufficient relevant information is available, slightly, to 469.

In empirically testing Equations (5.1)-(5.4), I focus the discussion on a majority subset of these farmers (331) but also confirm the results hold for all farmers (469).

The 331 observations on which I focus are the farmers who (i) believed at the outset they were purchasing Bt seeds in 2013, answering 'Yes' when asked pre-cultivation if they believed their seed was Bt effective, and who (ii) cultivated only one variety on the main plot. The first point roughly fixes for ex-ante expectations, in line with the the theoretical model, and excludes having to deal with the second largest group which answered 'I don't know', and which it is not clear can be considered a homogenous group. The second point allows me to exactly match the results from the biophysical test to the variety purchased; for farmers who cultivated more than one cotton variety on the plot from which the biophysical sample was taken in 2013 it is impossible to tell which variety the lab tests correspond to.

However, it is possible that adding farmers who did not believe they were planting Bt (43 farmers) and those who 'do not know' (73 farmers), and roughly fixing for these beliefs using dummy variables, can add information that alters the results; this would be a question of external validity. It is also possible that farmers who planted more than one variety (23 farmers across 53 observations) are naturally those who experiment more and learn better from their crop, which would make their exclusion in the sample bias the results downward.

To address these issues, for each of Equations (5.1)-(5.4), I include a column in the results that incorporates all 469 farmers across 499 observations.[19] I do so by incorporating all 'three' beliefs and using dummies to allow for differences in intercept and slope across belief, and by constructing a 'pseudo' Bt variable for farmers who cultivated more than one variety, based on the average Bt for that variety found for the other (one-variety per plot) farmers in the sample. The resulting back-in adding process incorporates all 469 farmers across 499 observations and serves as a check on the more focused results.

6 Results and discussion

6.1 Results

Table 2 shows the results from five versions of Equation (5.1). All are linear probability models to facilitate interpretation, with robust standard errors (adjusted for heteroskedasticity) including in subsequent tables. The 95% confidence intervals are noted below each coefficient.

[Table 2 here]

[19]The number of observations is greater than the number of farmers because of the farmers that cultivated more than one variety, with each variety counting as an observation.

Column 1 regresses *Change* only on standardized Bt level in 2013 and on district controls. Column 2 also takes into account variables that may be correlated with Bt level and impact the dependent variable: education, seed purchase price, and cotton selling price. As argued in the theoretical model, controlling for seed price is important because Bt content should be price-adjusted. Meanwhile, cotton selling price must be controlled for if it is correlated to Bt, i.e. if bollworms cause damage not only to yield but also to quality, which is captured in the cotton selling price variable.[20]

Column 3 adds variables which are exogenous to Bt level but may affect variety choice, whose inclusion therefore improves precision: farmer characteristics such as farming experience, planting history, and wealth. Column 4 adds the input variables whose role in the decision making process is more questionable. Farmers that intensify input use and obtain higher yield may be more inclined to keep the same variety the next year, or, behaving more rationally, they may distinguish that higher yield is due to own input choices thereby leaving variety choice unaffected. Yield is not included in the regressions because it qualifies as a bad control: the effect of Bt content, if learned, on farmer choice would operate largely through its effect on yield.

Whereas the above focus on the subset of 331 farmers, Column 5 runs the regression on all expanded 499 observations, by adding dummies for beliefs and interacting them with Bt level, and by constructing a 'pseudo' Bt measurement for farmers with more than on evariety.

The consistent result is that the Bt level as measured in-lab bears no effect on the proclivity to keep or change the seed variety in the next year. Point estimates are

[20]Cotton selling price is exogenous to each farmer's production since the farmers are small and therefore price takers.

very small and close to zero. They indicate that a one standard deviation increase in Bt level is associated with a change in the probability of variety change of 1.4% to 2.7%, depending on the specification, with no signifiance. A 95% confidence interval can rule out negative effects larger than 5% in absolute value across all specifications.

Regarding the control variables, the analysis confirms that higher cotton selling prices reduce the chance that the farmer will change the variety the next year, and this is almost significant at the 10% level. Seed purchase prices have coefficients that are significant but at levels very close to zero, confirming the qualitative evidence in the surveys that seed prices are neither high nor prohibitive in the Pakistani cotton market. The input coefficients are nearly all close to zero, implying that farmers who raise yield through input use realize that higher yield is due to input intensity and not necessarily seed quality, leaving variety choices unaffected.[21]

Farmer characteristics are evidently important: higher education and general farming experience increase the rate at which farmers change their varieties, suggesting that these farmers are more informed about different varieties and willing to experiment. Experience with Bt cotton cultivation and with the 2013 variety reduces the probability of variety change, suggesting that farmers become more comfortable with that variety over time and/or know how to cultivate it more efficiently, reducing the extent of variety change.

Table 3 shows that the results change significantly when we assess the impact of farmer perceptions of bollworm resistance on variety change. Column 1 regresses the dependent variable, *Change*, only on perceptions and district fixed effects, Column 2 adds farmer characteristics, seed purchase price, and cotton sale price as these can

[21]The exception is seeds sowed which is positive and significant. Possibly, varieties sowed more intensely were ones failing to grow properly, hence a higher likelihood that the variety is changed.

improve precision, and Column 3 adds two variables, pesticide use and yield, which may affect Change but whose exogeneity to perceptions is not clear.[22] Column 4 runs the last regression on the expanded set of 499 observations.

[Table 3 here]

The result across specifications is that farmers are less likely to change the variety purchased in 2014 when their perception of bollworm resistance for the 2013 season is more positive. Depending on the specification, farmers who viewed resistance as moderate are 11.1 to 15.5% less likely to change variety in the next year than those who viewed it as poor, and this is significant at the 10% or 5% level. Farmers who viewed resistance performance as very good are 17.0 to 19.5% less likely to change variety next year than those who viewed it as poor, and this is consistently significant at the 5% level. Therefore, farmers do change variety more often when they assess that the crop has exhibited poor resistance to bollworms. The controls possess similar signs and interpretations to those in **Table 2**.

Next, **Table 4** explores the role of Bt content in informing farmer perceptions. From the null result in **Table 2**, we would expect that farmers are unable to accurately assess the degree to which perceived resistance is an outcome of biophysical attributes (Bt), and this is corroborated.

[Table 4 here]

The dependent variable is perception of the farmer about bollworm resistance in 2013, lumped into Poor/Moderate or Very Good, and taking a binary value of 0 and 1 respectively, and the key independent variable is standardized Bt content. I

[22] Pesticide use may be driven by resistance perceptions, and yield and perceptions are likely correlated but it is not clear which affects which.

lump the dependent variable so I can perform a linear probability model, for ease of interpretation, but I perform robustness checks with an ordered logit (Section 7).

Column 1 regresses perception on standardized Bt content and on time and district controls (omitted). Column 2 adds education and seed purchase price because they may be correlated with Bt level and perceptions, while Column 3 adds years of experience and planting history to improve precision. Column 4 adds pesticide whose exogeneity is not clear: pesticide use may affect perceptions or existing perceptions may dictate pesticide use, or a combination of the two. Column 5 runs the analysis on all possible 499 observations.

In all specifications Bt content does not inform perception formation. The coefficients on standardized Bt content are small and insignificant, and positive effects greater than 2.7% can be ruled out in all specifications at the 95% level. The coefficients on farmer characteristics, planting history, and input use are also insignificant. Dummies on sowing time and district controls (omitted) are the only ones carrying some significance, indicating that perceptions are dictated largely by exogenous (time- and space- dependent) pest intensity or other unobservable or unmeasured factors.

Column 6 explores the possibility that perceptions of bollworm resistance are driven by yield outcomes. Yield cannot be included in the other regressions because it would be a bad control, so I regress the dependent perception variable only on log of yield per acre and on sowing time and district controls. The association is positive and significant: a 1% increase in yield per acre is associated with 14% greater likelihood of viewing resistance as very good instead of poor/moderate. This result is not causally identified: it is unclear whether yield informs perception or whether perception drives behavior that affects yield, since both responses were elicited from farmers during the same survey round. Nonetheless, this correlation result is robust including when inputs are controlled for.

Finally, **Table 5** examines the possibility that Bt content can be uncovered and impact not variety choice next season but pesticide use in the same season. Column 1 regresses pesticide use only on standardized Bt content as well as time of sowing, district, and soil-type controls (omitted). Column 2 adds education which may be correlated with Bt content and impact pesticide use, while Column 3 adds farmer characteristics, area cultivated, and the intensity of seeds planted per acre to improve precision. Column 4 incorporates irrigation and fertilizer use per acre since different inputs may be used in complementary quantities, though the direction of causation is not identified. Column 5 runs the second to last regression on the expanded set of observations.

Across specifications, Bt content does not significantly impact pesticide use. Other results are that more educated farmers use pesticide more, sowing seeds more intensively needs greater pesticide use, and fertilizer and pesticide use are complementary. In the final column, farmers who did not think they had Bt, or did not know, were less likely to use pesticide, suggesting that perhaps they did not believe their crop needed to be treated heavily for pests whether through biophysical properties or inputs.

[Table 5 here]

6.2 Discussion

The results are consistent and suggest that farmers are unable to learn about an important attribute of their seeds through cultivation experience, at least after one round of harvest. They are unable to distinguish the role of the seed itself in resistance (**Table 4**) and to switch varieties next year accordingly (**Table 2**). This is evidence of lack of learning, and not of unwillingness to switch, precisely because farmers do use switching to combat what they perceive as poor resistance (**Table 3**). Inability to discover Bt content through cultivation may also be evident in the absence of an appropriate response through input use (**Table 5**). The results correspond to row 4 in **Table 1**, with $\beta_1 < 0$, $\gamma_1 < 0$, $\theta_1 = 0$, $\phi_1 = 0$.

The absence of learning implies that market outcomes are stagnant. Average Bt content does not improve and farmers do not benefit from gradually enhanced varieties on the market. To calculate the extent of losses from lack of learning, I rely on informed estimates of the size of Pakistan's cotton cultivation industry and of the effect of Bt on damage abatement, detailed in **Appendix B**. Those estimates suggest that if average Bt improves in the long run from the in-sample level of 0.88 $\frac{\mu g}{g}$ to the maximum-effectiveness level of 1.59 $\frac{\mu g}{g}$, yield would have improved by up to \$170 million, or 12.5% of industry revenue in 2014.

Of course, these results are market and attribute-specific. Learning about an unknown attribute that has a clear effect on an observable outcome, because of the absence of confounding factors e_{it}, would be significantly easier. For example, in cotton, fiber whiteness is the outcome primarily of the seed's biophysical property, so white-fiber varieties could probably be deduced easily ex-post. Such examples notwithstanding, it is likely that many properties that are important for productivity are confounded by other factors, and are therefore difficult to deduce from cultivation experience alone, in the absence of certification standards or information provision.

7 Alternative explanations and robustness checks

The results from the following exercises are all presented in **Appendix D**.

7.1 Learning from others

The above assumes that if information is learned about Bt content, it is through the farmer's own cultivation experience. Neighbor effects are not included because for each farmer there are very few other farmers within the same village (smaller unit than district) who can therefore be possible peers. Also, there is no information about how far the villagers are, geographically or socially, so it is possible (even likely, with a stratified sample) that the farmers in each village are far apart and less relevant to each other than true next-door peers. This makes it very difficult to construct a measure of peer effects without a large degree of error and without introducing bias in the regression.

Nonetheless, as a rough attempt, I identify the other farmers in the same village as potential peers. If there is social learning, we expect a farmer to be more likely to switch the higher is the Bt of peers who purchased a different variety,[23] as well as expecting the own-Bt coefficient to become negative. The latter is because it is unclear how farmers can learn from their neighbors, if their neighbors cannot learn from their own experience.

Table D1 shows the results when such a peer variable (non-standardized) is integrated into the main regressions (of **Table 2**). Though in one specification the coefficient on this variable is significantly positive, it is not clear that this reflects true peer learning, given the high likelihood of measurement error in variable construction, the non-robustness of this significance in other specifications, and the fact that the coefficient on own-Bt remains null throughout.

[23]For most farmers there are almost no other villagers farming the same variety, making it impossible to construct a measure of Bt of peers who farmed the same variety, on which we would expect a negative coefficient with learning.

7.2 Different behavioral responses

It is possible that farmers react to low Bt content by switching suppliers in the next year instead of changing variety. I do not have data on supplier switching because suppliers are not named but I construct a best guess estimate for 207 observations (the others did not answer the questions necessary to construct this estimate). I assume the supplier changed if the farmer lists a different type of supplier institution in 2014 or if the farmer lists the same type of institution but the commuting time changed significantly. Based on this, I estimate that two-thirds of the farmers did not change their supplier. **Figure D1** shows no correlation between the change in supplier and Bt content. It appears unlikely that learning occurred and drove supplier switching.

Another possibility is that the farmers uncovered Bt content but reacted by exiting cotton production altogether. There is insufficient information to control for factors that influence exit, but, qualitatively, farmers who exited cite predominantly environmental reasons in the surveys as shown in **Figure D2**. Additionally, **Figure D3** shows no difference in the mean Bt gene expression between the group that exited and the one that remained.

7.3 Sample selection bias

The main tables check for sample selection bias by incorporating a column with all possible 499 observations for each exercise, in part by constructing a pseudo-Bt measure for farmers who cultivated more than one variety in 2013. **Table D2** shows that the results from the main regression do not change when the pseudo-Bt measure is used for all farmers,[24] including those who farmed one variety, for comparability.

[24] Selecting on the 'Yes' belief.

More broadly, **Table D3** shows that the farmers in the sample's focus group (331) and out (396), out of the total 737 farmers surveyed in Round 1.1 (but of whom only 501 finished all rounds), are similar in average age, years of farming experience, the area of the main plot they operate on, and the total area of land they own. The exception is education, with in-sample farmers having 0.6 years more of education on average, and this is significant at the 10% level. Since more educated farmers are more likely to learn from cultivation experience if learning is possible, this would push results in the sample to show more-than-average learning, but the results still demonstrate no learning.[25] Hence, the results from the in-sample regressions are likely representative, at least roughly, of farmers in the survey, who are in turn nationally representative, and the remaining differences would not drive the null effect.

7.4 Measurement error

It is possible that the behavioral models, empirical specifications, and sampling methods are sound, but that insignificance is due to attrition bias from measurement error in the key explanatory variable. The Bt variable is based on a sample of two random plants from each farmer's plot, taking a leaf and a boll from each plant. Whereas leaf values seem to be significantly correlated between the two plants for each farmer, the boll values seem to be much less correlated. Therefore, it is possible that that sample does not accurately represent the "true" Bt level of the farmer's variety (which can only be known by sampling all plants, destroying the plot).

In **Table D4** I redefine Bt content variable to reduce possible measurement error and rerun the regression in Column 3 in **Table 2**. In Column 1, I use an average of

[25] The other significant difference is province: the in-sample group is more heavily skewed toward Punjab than Sindh.

the leaf values only instead of leaf and boll. In Column 2, I still use the leaf values but with one value as an instrument for the other.[26] In Column 3, I use a subsample where the leaf values per farmer are nearly identical. As shown, the results do not change in any significant way. Therefore, while the size of the biophysical sample per farmer is small and measurement error may certainly exist, it does not appear that Bt content, even measured more restrictively, impacts seed choice as expected if learning is present.

Finally, one point of reassurance about the biophysical samples not being too far off mark are the findings in Ma et al (2017) that, in a damage abatement model, Bt content based off of the measurements 70 days after sowing significantly improves yield, holding all else fixed, for these farmers in the Pakistan Cotton Survey.

7.5 Additional robustness checks

To further check the robustness of the main regression, I focus on Column 3 in **Table 2** and introduce in **Table D5**: (i) a squared term for Bt level to allow for nonlinear effects, (ii) an interaction variable of Bt level with education to allow for differential effects by education level, and (iii) a variation where the variable "years that variety is grown" is a sequence of dummy variables, to allow for a nonlinear effect of cultivation years on variety choices. I also (iv) re-estimate the model with a logistic regression, using Firth's bias-reduced version of the logit which penalizes to prevent overfitting and small-sample bias. Therefore, this latter check in particular is very useful.

To check the effects of clustering the dependent variable in **Table 4**, I estimate

[26]The idea is that this will eliminate correlated noise or measurement error; a similar approach is used in Ashenfelter and Krueger (1994).

Column 3 as an ordered logit, with perceptions taking all three values and ordered as such (**Table D6**).

In all of these, the findings in Section 6 remain robust.

8 Conclusion

In developing countries, information challenges are ubiquitous and pronounced. It is often difficult to accurately evaluate financial borrowers, to design incentive-compatible mechanisms, to assess which technologies maximize efficiency, to know how to best adapt new technical and organizational skills, and to assess which government policies are most likely to support growth.

Agricultural producers in particular face rife information problems, including when they import and adapt foreign technologies for which local government certification and standardization are weak or nonexistent. With imported and adapted seed-based technologies, farmers are likely to not know important attributes if varieties are not certified, leaving room for potential learning ex-post by observing cultivation outcomes. In the absence of externally verifiable information and if heterogeneity of growing conditions mutes learning from peers, such a process of learning from own experience is particularly valuable. Learning about and plugging information gaps is not just a question of microeconomic behavior; at the macro level, if it allows farmers to make more informed choices over time then it improves productivity, with implications for growth and competitiveness.

Drawing on this context, I model a process whereby farmers can learn about the variety through cultivation experience and make more informed decisions in the next season. I then use this model to derive econometric specifications to test for learning, since whether agents can learn and redress information problems is ultimately an empirical question. Using a rich dataset, I apply the empirical exercises to cotton cultivation in Pakistan, where there is imperfect information about an imported and adapted pest-resistance technology (the Bt gene). I use a number of behavioral outcomes to evaluate whether farmers can learn about this attribute of their seed on which they lack prior information.

The results indicate that cultivation experience is not sufficient to redress the information gap. Farmers are unable to uncover the Bt content of their crop even after cultivation and harvest are complete, likely because of the existence of other confounding factors that are difficult to measure or parse out. As a result, Bt content does not inform farmer perceptions of their crop's pest resistance nor their choices about variety purchases in the next season. This impedes gains at the farmer level as well as wider improvement in crop productivity in the Pakistani cotton market. The absence of learning is robust across different specifications and behavioral outcomes that can signal learning, and points to a persistent information failure in the absence of external policy intervention.

Nonetheless, the prescription of external information provision as a solution is qualified. In the case of the Pakistani cotton market, information provision by the government is itself difficult, given that the weak capacities of the Pakistani state contributed to the proliferation of information failures in the first place. Therefore, the paper illustrates the dual dilemma in many developing countries, where market failures must be addressed by potentially equally limited government institutions. Any policy solution to address information failures must take both private and public constraints into account to be effective.

FIGURES

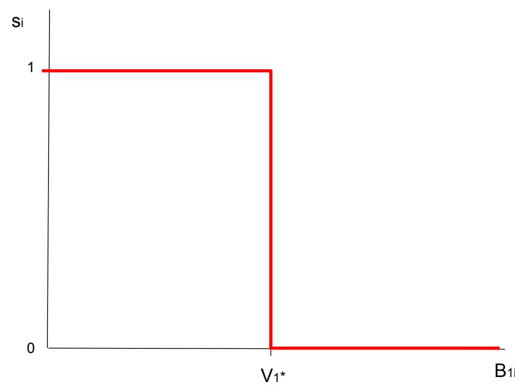

Figure 1: Discrete switching

Figure 1 shows whether or not producers switch varieties next year if learning about Bt content is possible. Those who find out Bt content exceeded their expectations do not switch and vice versa.

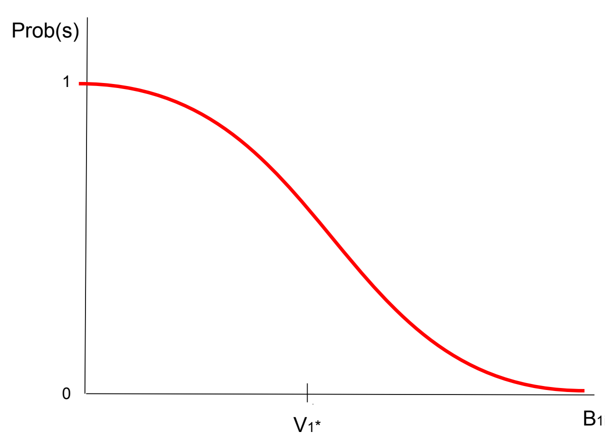

Figure 2: Probabilistic switching

Figure 2 shows switching as a smooth function of Bt content. Learning occurs but is easier at the extremes; farmers are more likely to switch the further below expectations Bt content is.

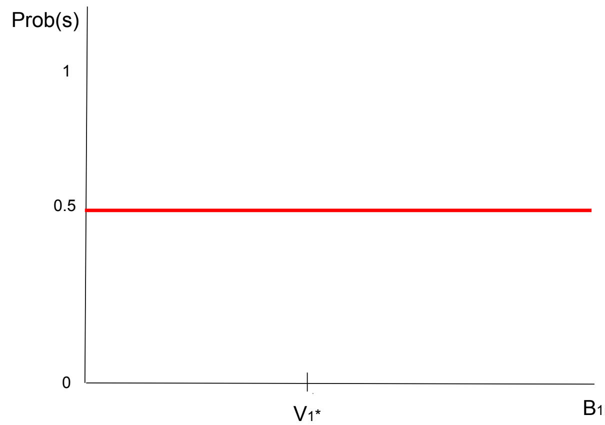

Figure 3: No learning

Figure 3 shows that if it is impossible to deduce Bt, the probability of switching is constant for all values (here, a tossup).

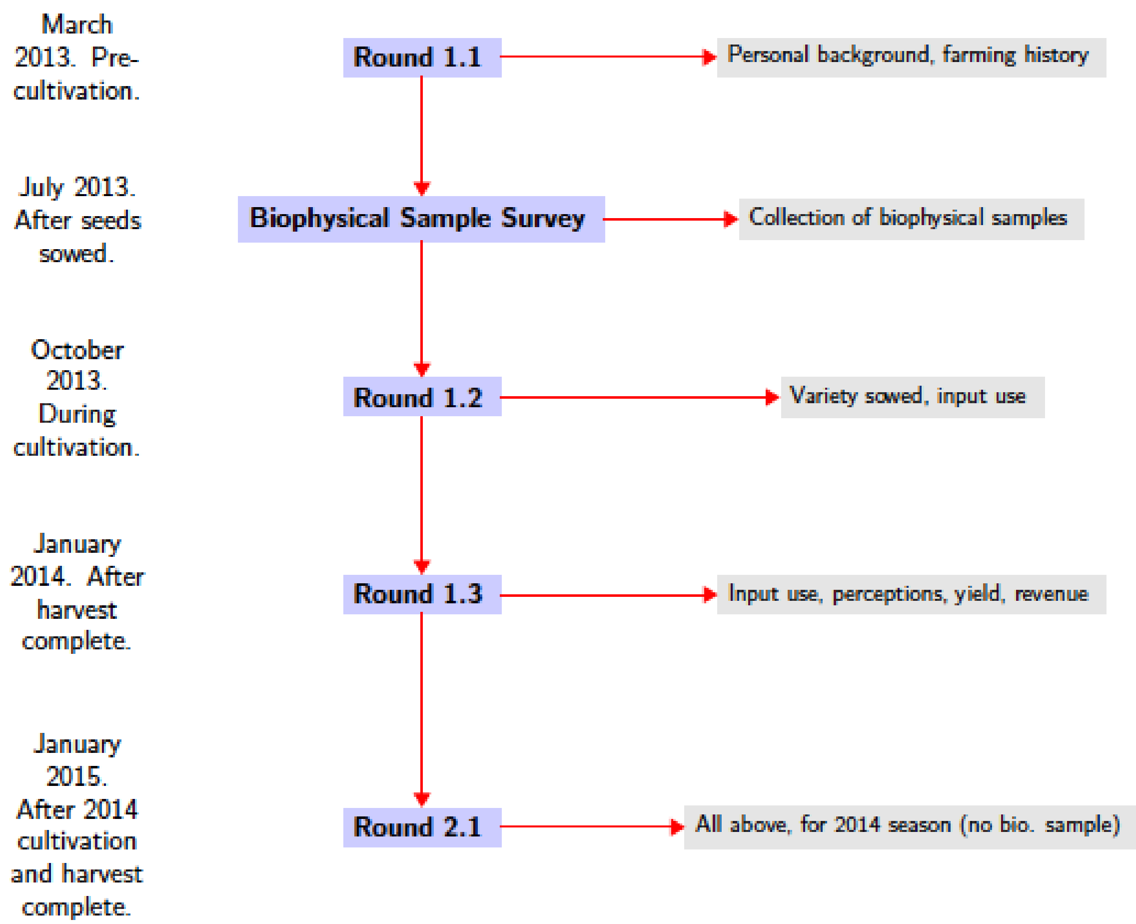

Figure 4: Timeline of surveys and cultivation

Figure 4 describes the structure of the Pakistan Cotton Survey, chronologically and content-wise. For each survey, I note the date it was taken, its title, and some of the pertinent questions asked.

TABLES

Table 1: Coefficient combinations and interpretation

β_1	γ_1	θ_1	ϕ_1	**Interpretation**
<0	<0	>0		Farmers learn and change variety accordingly.
$= 0$	$= 0$	> 0		Farmers learn but do not change variety.
$= 0$	$= 0$	> 0	< 0	Farmers learn but respond by changing input.
$= 0$	< 0	$= 0$	$= 0$	Farmers are unable to learn.

Table 1 summarizes the possible meaningful combinations of coefficients and their corresponding economic interpretations. The coefficients in Columns 1-4 are derived from Equations (5.1)-(5.4). β_1 measures the effect of an increase of Bt level on the probability of variety change; γ_1 measures the effect of improved perceptions of bollworm resistance on the probability of variety change; θ_1 measures the effect of an increase of Bt level on the probability of having improved perceptions; and ϕ_1 measures the effect of an increase of Bt level on the use of pesticides.

Table 2: Effect of Bt on variety change

	Dependent variable:				
	CHANGED				
	(1)	(2)	(3)	(4)	(5)
Bt (standardized)	0.015	0.014	0.015	0.015	0.027
	(−0.048, 0.078)	(−0.049, 0.077)	(−0.048, 0.078)	(−0.049, 0.079)	(−0.034, 0.087)
Belief: No					0.111
					(−0.076, 0.298)
Belief: Don't know					0.030
					(−0.106, 0.167)
Education		0.006	0.014**	0.013**	0.014***
		(−0.005, 0.017)	(0.002, 0.025)	(0.001, 0.025)	(0.004, 0.024)
Farming Experience			0.006**	0.006**	0.005**
			(0.001, 0.012)	(0.00005, 0.011)	(0.0004, 0.009)
Yrs grown variety			−0.082***	−0.084***	−0.062***
			(−0.132, −0.031)	(−0.136, −0.032)	(−0.103, −0.020)
Yrs grown Bt			−0.040**	−0.037**	−0.011
			(−0.072, −0.008)	(−0.069, −0.005)	(−0.039, 0.017)
Land owned			−0.008**	−0.008***	−0.007**
			(−0.014, −0.002)	(−0.014, −0.002)	(−0.013, −0.001)
Seed price		−0.0004	−0.001*	−0.0005*	−0.0004*
		(−0.001, 0.0001)	(−0.001, 0.00001)	(−0.001, 0.0001)	(−0.001, 0.00001)
Cotton selling price		−0.020	−0.022	−0.024	−0.013
		(−0.048, 0.008)	(−0.050, 0.007)	(−0.053, 0.005)	(−0.036, 0.010)
Irrigation				−0.0001*	
				(−0.0002, 0.00001)	
Fertilizer				−0.001	
				(−0.002, 0.001)	
Seed amount				0.025**	
				(0.001, 0.049)	
Labor				0.0002	
				(−0.0005, 0.001)	
Pesticide				−0.009	
				(−0.048, 0.030)	
Bt*Belief:No					−0.034
					(−0.231, 0.163)
Bt*Belief:Don't know					0.022
					(−0.092, 0.137)
District FE	Yes	Yes	Yes	Yes	Yes
Observations	331	331	331	331	499
R^2	0.206	0.218	0.279	0.299	0.211
Adjusted R^2	0.143	0.148	0.205	0.213	0.146
Residual Std. Error	0.462 (df = 306)	0.461 (df = 303)	0.445 (df = 299)	0.443 (df = 294)	0.460 (df = 460)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2 demonstrates the results of Equation (5.1). Across specifications, Bt content does not influence variety change next year.

Table 3: Effect of perceptions on variety change

	Dependent variable:			
	CHANGED			
	(1)	(2)	(3)	(4)
Moderate	−0.137*	−0.155**	−0.137*	−0.111
	(−0.288, 0.015)	(−0.304, −0.006)	(−0.290, 0.016)	(−0.260, 0.037)
VeryGood	−0.195**	−0.193**	−0.170**	−0.172**
	(−0.357, −0.033)	(−0.350, −0.037)	(−0.330, −0.009)	(−0.327, −0.018)
Belief: No				0.052
				(−0.337, 0.442)
Belief: Don't know				−0.086
				(−0.352, 0.180)
Education		0.015**	0.015**	0.015***
		(0.003, 0.026)	(0.004, 0.027)	(0.005, 0.025)
Farming Experience		0.007**	0.007**	0.005**
		(0.001, 0.012)	(0.001, 0.012)	(0.001, 0.009)
Yrs grown variety		−0.077***	−0.080***	−0.060***
		(−0.127, −0.027)	(−0.131, −0.029)	(−0.101, −0.019)
Yrs grown Bt		−0.042**	−0.043***	−0.009
		(−0.074, −0.010)	(−0.075, −0.010)	(−0.038, 0.019)
Land owned		−0.008***	−0.008***	−0.007**
		(−0.014, −0.002)	(−0.014, −0.002)	(−0.013, −0.001)
Seed price		−0.001**	−0.001**	−0.0003*
		(−0.001, −0.00002)	(−0.001, −0.00000)	(−0.001, 0.0001)
Cotton selling price		−0.020	−0.019	−0.011
		(−0.049, 0.010)	(−0.048, 0.011)	(−0.035, 0.012)
Pesticide			−0.006	−0.0002
			(−0.040, 0.028)	(−0.011, 0.011)
Log yield			−0.058	−0.027
			(−0.153, 0.037)	(−0.109, 0.055)
Moderate*Belief:No				0.048
				(−0.356, 0.453)
Moderate*Belief:Don't know				0.162
				(−0.153, 0.477)
VeryGood*Belief:No				0.123
				(−0.342, 0.588)
VeryGood*Belief:Don't know				0.075
				(−0.260, 0.410)
District FE	Yes	Yes	Yes	Yes
Observations	331	331	331	499
R^2	0.220	0.294	0.297	0.220
Adjusted R^2	0.156	0.218	0.216	0.147
Residual Std. Error	0.459 (df = 305)	0.442 (df = 298)	0.442 (df = 296)	0.460 (df = 455)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3 demonstrates the results of Equation (5.2). Across specifications, improved perceptions of bollworm resistance decrease the probability that the farmer will change varieties next year.

Table 4: Effect of Bt on farmer perceptions

	Dependent variable:					
	VeryGood					
	(1)	(2)	(3)	(4)	(5)	(6)
Bt (standardized)	−0.026	−0.027	−0.027	−0.028	−0.035	
	(−0.079, 0.027)	(−0.080, 0.026)	(−0.080, 0.025)	(−0.081, 0.025)	(−0.081, 0.012)	
Belief: No					−0.067	
					(−0.246, 0.111)	
Belief: Don't know					−0.107	
					(−0.236, 0.022)	
Education		0.003	0.002	0.002	−0.002	
		(−0.009, 0.015)	(−0.010, 0.015)	(−0.010, 0.015)	(−0.011, 0.007)	
Farming Experience			0.001	0.001	0.003	
			(−0.005, 0.007)	(−0.004, 0.007)	(−0.001, 0.008)	
Yrs grown variety			0.013	0.012	−0.009	
			(−0.041, 0.066)	(−0.041, 0.066)	(−0.051, 0.034)	
Yrs grown Bt			0.011	0.010	−0.006	
			(−0.024, 0.046)	(−0.025, 0.045)	(−0.032, 0.021)	
Seed price		0.0001	0.0001	0.0001	0.0002	
		(−0.0004, 0.001)	(−0.0004, 0.001)	(−0.0004, 0.001)	(−0.0001, 0.001)	
Fertilizer			−0.0001	−0.00003	0.00003	
			(−0.002, 0.001)	(−0.002, 0.002)	(−0.0001, 0.0002)	
Pesticide				−0.004	0.007*	
				(−0.048, 0.040)	(−0.001, 0.016)	
Log yield						0.140***
						(0.051, 0.229)
Bt*Belief:No					−0.005	
					(−0.148, 0.138)	
Bt*Belief: Don't know					0.011	
					(−0.065, 0.087)	
District and sowing time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	331	331	331	331	499	331
R^2	0.276	0.277	0.279	0.279	0.301	0.293
Adjusted R^2	0.190	0.185	0.177	0.174	0.224	0.209
Residual Std. Error	0.447 (df = 295)	0.448 (df = 293)	0.451 (df = 289)	0.451 (df = 288)	0.437 (df = 449)	0.441 (df = 295)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4 demonstrates the results of Equation (5.3). In Columns 1-5, Bt content does not systematically influence farmer perceptions of the bollworm-resistance performance of their crop. Column 6 shows that yield is positively correlated with farmer perceptions.

Table 5: Effect of Bt on pesticide use

	Dependent variable:				
	Pesticide use				
	(1)	(2)	(3)	(4)	(5)
Bt (standardized)	−0.039	−0.045	−0.026	−0.014	−0.040
	(−0.273, 0.195)	(−0.277, 0.186)	(−0.259, 0.207)	(−0.220, 0.191)	(−0.372, 0.292)
Belief: No					−1.343**
					(−2.425, −0.261)
Belief: Don't know					−1.234***
					(−2.006, −0.463)
Education		0.032	0.037*	0.029	0.075
		(−0.009, 0.074)	(−0.006, 0.080)	(−0.015, 0.072)	(−0.027, 0.178)
Farming Experience			0.008	0.009	0.0003
			(−0.010, 0.026)	(−0.008, 0.026)	(−0.038, 0.039)
Yrs grown variety			−0.073	−0.089	−0.251
			(−0.214, 0.069)	(−0.224, 0.047)	(−0.596, 0.093)
Yrs grown Bt			−0.094	−0.085	0.415**
			(−0.209, 0.020)	(−0.192, 0.021)	(0.069, 0.761)
Land owned			−0.0002	0.006	0.004
			(−0.018, 0.017)	(−0.010, 0.023)	(−0.045, 0.053)
Area cultivated			0.004	0.001	−0.015
			(−0.016, 0.023)	(−0.018, 0.021)	(−0.041, 0.010)
Irrigation				0.0001	
				(−0.0002, 0.0004)	
Fertilizer				0.013***	
				(0.007, 0.018)	
Seed amount			0.101**	0.101***	−0.018
			(0.023, 0.179)	(0.032, 0.170)	(−0.167, 0.130)
Bt*Belief: No					0.299
					(−0.269, 0.866)
Bt*Belief: Don't know					0.361
					(−0.117, 0.839)
District, sowing time, and soil FE	Yes	Yes	Yes	Yes	Yes
Observations	331	331	331	331	499
R^2	0.342	0.350	0.375	0.438	0.236
Adjusted R^2	0.262	0.268	0.281	0.349	0.151
Residual Std. Error	1.392 (df = 294)	1.386 (df = 293)	1.374 (df = 287)	1.307 (df = 285)	4.202 (df = 448)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5 demonstrates the results of Equation (5.4). Across specifications, Bt content does not affect pesticide use.

References

[1] Ali, A and Abdulai, A. (2010) The adoption of genetically modified cotton and poverty reduction in Pakistan. *Journal of Agricultural Economics,* 61:175-192.

[2] Ali S, Shahid H, Shahid M, Ali GM, and Yusu Z. (2010) Status of Bt cotton cultivation in major growing areas of Pakistan. *Pakistan Journal of Botany,* 42:1583-1594.

[3] Ali S, Shah SH, Ali GM, Iqbal A, Asad M, and Zafar Y. (2012) Bt Cry toxin expression profile in selected Pakistani cotton genotypes. *Current Science,* 102:1632-1636.

[4] Ashenfelter O and Krueger A. (1994) Estimates of the economic return to schooling from a new sample of twins." *American Economic Review*, 84(5):1157-1173.

[5] Auriol E and Schilizzi S. (2015) Quality signaling through certification in developing countries *Journal of Development Economics,* 116:105-121.

[6] Berazneva J, McBride L, Sheahan M, and Guerena D. (2016) Perceived, measured, and estimated soil fertility in east Africa. AAEA 2016 Annual Meeting.

[7] Bold T, Kaizzi K, Yanagizawa-Drott D, and Svensson J. (2017) Lemon technologies and adoption: Measurement, theory and evidence from agricultural markets in Uganda. *Quarterly Journal of Economics*, 132:1055-1100.

[8] Braverman A (1985). Rural credit markets and institutions in developing countries. *World Development,* 14(10):1253-1267.

[9] Conley T and Udry C. (2010) Learning about a new technology: Pineapple in Ghana. *American Economic Review,* 100:35-69.

[10] Crane-Droesch A. (2018) Technology diffusion, outcome variability, and social learning: evidence from a field experiment in Kenya. *American Journal of Agricultural Economics,* 100:955-974.

[11] Darby M and Karni E. (1973) Free competition and the optimal amount of fraud. *The Journal of Law and Economics,* 16:67-88.

[12] Dean M. (2019) Rational inattention lecture. Department of Economics, Columbia University, personal website.

[13] Dulleck U and Kerschbamer R. (2006) On doctors, mechanics, and computer specialists: the economics of credence goods. *Journal of Economic Literature,* 44:5-42.

[14] Dulleck U, Kerschbamer R, and Sutter M. (2011) The economics of credence goods: An experiment on the role of liability, verifiability, reputation and competition. *American Economic Review,* 101:526-55.

[15] FAO (2009). Responding to the challenges of a changing world. *Proceedings of the Second World Seed Conference,* Rome Italy, September 8-10..

[16] Foster, A. and Rosenzweig M. (1995) Learning by doing and learning from others: human capital and technical change in agriculture. *Journal of Political Economy,* 103(6):1176-1209.

[17] Foster, A. and Rosenzweig M. (2010) Microeconomics of technology adoption. *Annual Review of Economics,* 2:395-424.

[18] Ghosh P, Mookherjee D, and Ray D. (2000) Credit rationing in developing countries: An overview of the theory. In Mookherjee D and Ray D, eds., *A Reader in Development Economics.* London: Blackwell.

[19] Girard T and Dion P. (2010) Validating the search, experience, and credence product classification framework. *Journal of Business Research,* 63:1079-1087.

[20] Hanna R, Mullainathan S, and Schwartzstein J. (2014) Learning through noticing: Theory and evidence from a field experiment. *Quarterly Journal of Economics,* 129:1311-1353.

[21] International Food Policy Research Institute (IFPRI); Innovative Development Strategies (Pvt) Ltd.; National Institute of Genomics and Advanced Biotechnology (NIGAB), National Agricultural Research Center (NARC); Office of Research, Innovation, and Commercialization (ORIC), University of Agriculture Faisalabad (UAF); Department of Agricultural Economics, University of Agriculture Faisalabad (UAF) (2017). *Pakistan Cotton Survey.* Harvard Dataverse.

[22] ISAAA. (2017) Global status of commercialized Biotech/GM crops in 2017. *ISAAA Brief* No. 53.

[23] Khan, M. (2010) "Learning, technology acquisition, and governance challenges in developing countries." Unpublished manuscript, SOAS Economics Dpt.

[24] Kouser S and Qaim M. (2014) Bt cotton, damage control and optimal levels of pesticide use in Pakistan. *Environment and Development Economics,* 19:704-723.

[25] Ma X, Smale M, Spielman DJ, Zambrano P, Nazli H, and Zaidi F. (2017) A question of integrity: variants of Bt cotton, pesticides, and productivity in Pakistan. *Journal of Agricultural Economics,* 68:366-385.

[26] Ma X and Shi G. (2015) A dynamic adoption model with Bayesian learning: an application to US soybean farmers. *Agricultural Economics,* 46:25-38.

[27] Maertens A, Michelson H, and Nourani V. (2018) How do farmers learn from extension services? Evidence from Malawi'. *SSRN Electronic Journal.*

[28] Munshi K. (2004) Social learning in a heterogenous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics,* 73:185-213.

[29] Murphy D. (2017) Underground knowledge: soil testing, farmer learning, and input demand in Kenya. Unpublished.

[30] Pemsl D, Waibel H, and Gutierrez A. (2005) Why do some Bt cotton farmers in China continue to use high levels of pesticides? *International Journal of Agricultural Sustainability,* 3:44-56.

[31] Poole, N. (2017) *Smallholder Agriculture and Market Participation.* Rugby, UK: Practical Action Publishing.

[32] Sims C. (2003) Implications of rational inattention. *Journal of Monetary Economics,* 50:665-690.

[33] Spielman DJ, Nazli H, Ma X, Zombrano P and Zaidi F. (2015) Technological opportunity, regulatory uncertainty, and the economics of Bt cotton in Pakistan. *The Journal of Agrobiology Management and Technology,* 18:98-112.

[34] Spielman DJ, Zaidi F, Zambrano P, Khan A, Ali S, Cheema H, Nazli H, Khan R, Iqbal A, Zia M, and Ali G. (2017) What are farmers really planting? Measuring the presence and effectiveness of Bt cotton in Pakistan. *PLOS ONE* Vol. 12.

[35] Stiglitz, J. (1989). Markets, market failures, and development. *American Economic Review.* 79:197-203.

[36] Stiglitz J, Udry C, Siamwalla A, Bell C and Aleem I (1990). A symposium issue on imperfect information and rural credit markets. *World Bank Economic Review.*

[37] Tjernstrom E. (2014) Signals, similarity and seeds: Social learning in the presence of imperfect information and heterogeneity. Unpublished.

[38] USDA Foreign Agricultural Services (2019). Pakistan: Cotton and Products Annual.

[39] Waldman K, Kerr J, and Isaacs K. (2014) Combining participatory crop trials and experimental auctions to estimate farmer preferences for improved common bean in Rwanda. *Food Policy.* 46:183-192.

Appendix A: Theoretical model details

The expected value of Bt in the next period for each farmer is the expected outcome for switching or not, Equations (1.1)-(1.2), weighted by that probability:

$$\begin{aligned} E_i(\tilde{B_{2i}}) &= Prob(S)\left[E(B_1)\right] + (1 - Prob(S))\left[p\tilde{B_{1i}} + (1-p)E(B_1)\right] \\ &= E(B_1)\left[1 - p(1 - Prob(S))\right] + p(1 - Prob(S))\tilde{B} \end{aligned} \tag{A1}$$

To find the expected change in outcome from $t = 1$ to $t = 2$ for a farmer with initial realization $\tilde{B_{1i}}$, we subtract $\tilde{B_{1i}}$ from (A1):

$$\begin{aligned} E_i(\Delta\tilde{B_i}) &= E_i(\tilde{B_{2i}}) - \tilde{B_{1i}} \\ &= E(B_1)\left[1 - p(1 - Prob(S))\right] + p(1 - Prob(S))\tilde{B_{1i}} - \tilde{B_{1i}} \\ &= \left(E(B_1) - \tilde{B_{1i}}\right)\left[1 - p(1 - g(\tilde{B_{1i}} - V_1^*))\right] \end{aligned} \tag{A2}$$

Expected change in Bt content across the market is found by taking expected change for each initial realization $\tilde{B_{1i}}$, (A2), weighing it by the probability of its occurrence in the first period $Prob(\tilde{B_{1i}})$, and summing across:

$$\begin{aligned} E(\Delta\tilde{B}) &= \sum_i Prob(\tilde{B_{1i}}) * E_i(\Delta\tilde{B_i}) \\ &= \sum_i Prob(\tilde{B_{1i}}) * \left(E(B_1) - \tilde{B_{1i}}\right)\left[1 - p(1 - g(\tilde{B_{1i}} - V_1^*))\right] \end{aligned} \tag{A3}$$

A simpler expression can be obtained for (A3) - the expected change in Bt averaged across all farmers - which also allows us to see how it is affected by various parameters. To see that, assume g takes a specific functional form: a linear form $g = -\alpha(\tilde{B_{1i}} - V_1^*) + 0.5$, where $\alpha > 0$; the parameter α is the *learning parameter*. This form guarantees that for values at the expectation switching is a tossup,

$g(0) = 0.5$, though having any constant g_0 instead of 0.5 does not affect any results. Furthermore, allowing for a discrepancy between expectations and true market averages, $V_1^* = E(B_1) + \gamma$. Substituting into (A3):

$$
\begin{aligned}
E(\Delta\tilde{B}) &= \sum_i Prob(\tilde{B_{1i}}) * \left(E(B_1)^* - \tilde{B_{1i}}\right) \left[1 - p(1 + \alpha\{\tilde{B_{1i}} - [E(B_1) + \gamma]\} - 0.5)\right] \\
&= \sum_i Prob(\tilde{B_{1i}}) * \left(E(B_1) - \tilde{B_{1i}}\right) \left[1 - 0.5p + \alpha p(E(B_1) + \gamma - \tilde{B_{1i}})\right]
\end{aligned}
\tag{A4}
$$

(A4) can be written in continuous form. Replacing the summation with integration, and letting $f(\tilde{B_{1i}})$ be the probability density function, then:

$$
E(\Delta\tilde{B}) = \int f(\tilde{B_{1i}}) \left(E(B_1) - \tilde{B_{1i}}\right) \left[1 - 0.5p + \alpha p(E(B_1) + \gamma - \tilde{B_{1i}})\right] d\tilde{B_{1i}} \tag{A5}
$$

When simplified, (A5) reduces to a very simple expression. To see that:

$$
\begin{aligned}
E(\Delta x) &= \int f(x)\,(E(x)-x)\,[1-0.5p+ap(E(x)+\gamma-x)]\,dx \\
&= \int f(x)(E(x)-x)dx - 0.5p\int f(x)(E(x)-x)dx \\
&+ \alpha p\int f(x)(E(x)-x)(E(x)+\gamma-x)dx \\
&= 0+0+\alpha p\int f(x)(E(x)-x)(E(x)+\gamma-x)dx \\
&= \alpha p\int f(x)(E(x)-x)(E(x)-x)dx + \gamma\alpha p\int f(x)(E(x)-x)dx \\
&= \alpha p\int f(x)\left[(E(x))^2-2E(x)x+x^2\right]dx+0 \\
&= \alpha p\left[(E(x))^2\int f(x)dx - 2E(x)\int f(x)xdx + \int f(x)x^2dx\right] \\
&= \alpha p\left[(E(x))^2(1) - 2E(x)E(x) + \int f(x)x^2dx\right] \\
&= \alpha p\left[-(E(x))^2 + \int f(x)x^2dx\right] \\
&= \alpha p\left[-E(x)^2+E(x^2)\right] \\
&= \alpha p\left[E(x^2)-(E(x))^2\right] \\
&= \alpha p Var(x) \geq 0
\end{aligned}
$$

Therefore, by assuming a linear form for g, and since $Var(\tilde{B}_1)=\sigma^2$, we get:

$$E(\Delta\tilde{B}) = \alpha p\sigma^2 \tag{A6}$$

Appendix B: Calculating benefits to farmers

Since Bt imbues resistance to bollworms and improves cotton yield, higher average Bt quality on the market due to experience-based learning and selection should improve overall industry performance. How would Bt content improvement translate into monetary gains for the farmers?

This estimation proceeds in three parts. First, I estimate the size of Pakistan's cotton cultivation industry in the 2014-2015 season, the year for which I test learning and heuristic response by farmers. Second, I estimate the effect of varying levels of Bt on cotton yield and revenue. Third, I calculate changes in Bt content for different learning parameters α - with varying ranges of p and σ - and apply the results in Steps 1-2 to derive monetary benefits to farmers.

1. Calculating the size of the industry is straightforward. According to the Pakistani government, cotton production in Pakistan in 2014-2015 totaled 13,960,000 bales, equivalent to about 2.37 billion kg. From my data, the average price, in Pakistani Rupees, that farmers received for their 2014 crop per 40 kg mound of cotton was about 2313 PR, or 23 USD. Since 2.37 billion kg is equivalent to 59.3 million (40 kg) mounds, multiplying that amount by the price received per mound totals 1.364 billion USD, or 0.5% of the country's GDP for that year. Of course, this is only what the farmers receive - there is more value added downstream.

2. Calculating the effect of Bt improvement on yield and revenue is more complicated. Ma et al (2016) suggest the following breakdown of lethality:

Table B1: Bt content and pest lethality

Bt content ($\mu g/g$)	Lethal level (% pests killed)
0.60	50
0.70	60
0.88	70
1.06	80
1.34	90
1.59	95

Table B1 can be used to extrapolate differences in lethality based on Bt content. For example, an improvement in mean Bt content from 0.88 $\frac{\mu g}{g}$ to 0.97 $\frac{\mu g}{g}$ would raise killing effectiveness from 70% to 75%. The question is how this corresponds to output gain. Research suggests that Bt can protects half of all yield from destruction; if a maximum lethal level of 100% effective Bt improves yield by 50%, then 5% increase in lethal levels improves yield by 2.5%, or, given the size of the Pakistani cotton cultivation industry, about 34 million USD.

3. In this way, we can translate estimated improvement in average Bt content in the next year, $E(\Delta\tilde{B}) = \alpha p \sigma$, into monetary gains. For example, **Figure B1** below shows monetary gains, on the vertical axis, in millions of USD for $\alpha \in [0, 0.2]$, $p \in [0, 1]$, $\sigma = 0.57$ (this is the standard deviation in my data). Take a specific point such as $\alpha = 0.1$ and $p = 0.8$. $E(\Delta\tilde{B}) = 0.1 * 0.8 * 0.57 = 4.56$, so that average Bt shifts from 0.88 to 0.926 $\frac{\mu g}{g}$. In turn, using the methodology in Steps 1-2, this would generate monetary gains of about 17.4 million dollars, shaded in bluish green on the figure. With higher σ the graph would tilt further up, generating more gains for any combination of learning and variety integrity.

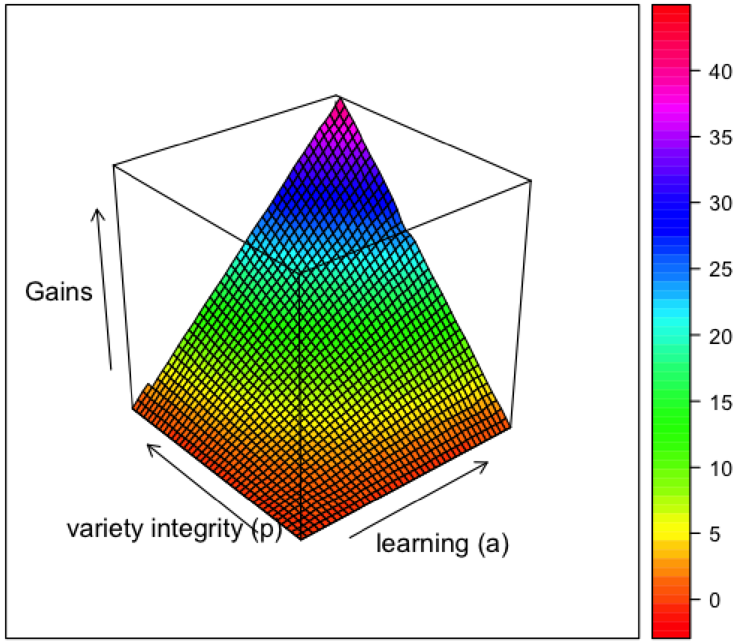

Figure B1: Illustrating monetary gains for farmers

Figure B1 plots an example of market-wide monetary gains in one year in millions of USD as a function of both the learning parameter (axis from 0-0.20) and variety integrity (axis from 0-1), with σ=0.57.

This method, though based on the short run, also provides a rough back-of-the-envelope estimate of maximum possible gains from learning in the long-run. If Bt improves in the long run from the in-sample level of 0.88 $\frac{\mu g}{g}$ to the maximum-effectiveness level of 1.59 $\frac{\mu g}{g}$, this 25% improvement in percent of lethal pests killed results in 12.5% improvement in yield (According to Step 2), and therefore gains of up to 170 million USD in 2014. Actual long term gains would of course depend on shifts in relative prices between less and more effective varieties.

Appendix C: Variable construction and distribution

The personal, price, and input controls are constructed as follows.

Education is the number of years of schooling of the household head by 2013. *Farming experience* is the years of general farming experience of the head by 2013. *Years Bt grown* is the total number of years that the household has grown (what they think are) Bt varieties, including and up to 2013. *Years variety grown* is the number of years in total that the farmer has grown the specific 2013 variety, including and up to 2013. *Land owned* is the amount of land, in acres, owned by the household in 2012.

Seed purchase price is the price, in 2013 Pakistani rupees, at which the farmer purchased one kilogram of seeds of the target variety in 2013. *Selling price* is the price, in 2013 hundreds of Pakistani rupees, at which the farmer sold one 40 kilogram mound of the variety cultivated and harvested in 2013.

Irrigation is a measure of the total minutes of irrigation per acre of cotton cultivated in 2013. *Fertilizer* measures the extent of nitrogen-fertilizer used, as kilograms per acre of cotton cultivated in 2013. I calculate it by multiplying the nitrogen percent of each type of fertilizer with the amount (in kg) used. *Seed amount* is the amount of seeds sowed for that variety in kilograms per acre of cotton cultivated in 2013. *Labor* measures the total number of labor hours that were reported worked, per acre, during the 2013 season.

Table C1 provides a summary of the distribution of key variables in the data, including the dependent variables. Variety change between 2013 and 2014 occurred in 55.8% of the sample. The average level of Bt expression is 0.877 micrograms of *cry* protein per gram of plant tissue. This is only moderately high: a measurement of 0.598 means the plant has 50% chance of killing bollworms at specific conditions while a level of 1.59 offers a 95% chance of doing so. Therefore, on average, the farmers are not cultivating very effective Bt varieties. [27]

The table also shows that the average farmer sampled has 5 years of education, 16 years of general farming experience, 4 years of experience cultivating Bt varieties, and has cultivated the 2013 variety for 2 years (including 2013); owns 6.5 acres of agricultural land; purchased seeds for about 280 Pakistani rupees ($1.80) per kilogram of seeds and sold the crop at 2,700 rupees ($17.30) per 40 kg mound of cotton; irrigated each acre cultivated for 23 hours total; applied 85 kilograms of fertilizer and 2.4 liters of pesticide per acre cultivated; sowed 7 kilograms of seeds per acre; and put in 163 hours of labor total per acre.

The histograms in **Figure C1** illustrate these distributions.

Table C1: Distribution of Variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Changed	331	0.538	0.499	0	0	1	1
Bt level ($\mu g/g$)	331	0.88	0.57	0.00	0.48	1.14	3.50
Education	331	5.0	4.8	0	0	9	20
Farming experience	331	16.1	10.7	2	7	22	49
Years variety grown	331	2.1	1.1	1	1	3	7
Years Bt grown	331	4.2	1.6	1	3	5	11
Land owned (acres)	331	6.7	9.3	0.0	2.0	8.0	67.0
Seed price (PR)	331	289.4	126.8	100.0	200.0	350 .0	900.0
Selling price ('00 PR)	331	27.7	2.6	18.0	26.4	29.7	34.0
Irrigation (mts/acre)	331	1,388	776	120	810	1,835	4,620
Fertilizer (kg/acre)	331	85.62	36.77	0	59.80	103.00	236.00
Seed amount (kg/acre)	331	6.94	2.81	2.00	5.00	9.00	16.00
Pesticide (L/acre)	331	2.41	1.62	0.00	1.30	3.20	10.00
Labor (hours/acre)	331	163.4	93.7	36.0	103.6	204.6	500.0

Table C1 summarizes the distribution of the key variables used in the analysis.

[27] With regard to perception, 18% of farmers rated the bollworm resistance performance as poor, 40% as moderate, and 42% as very good.

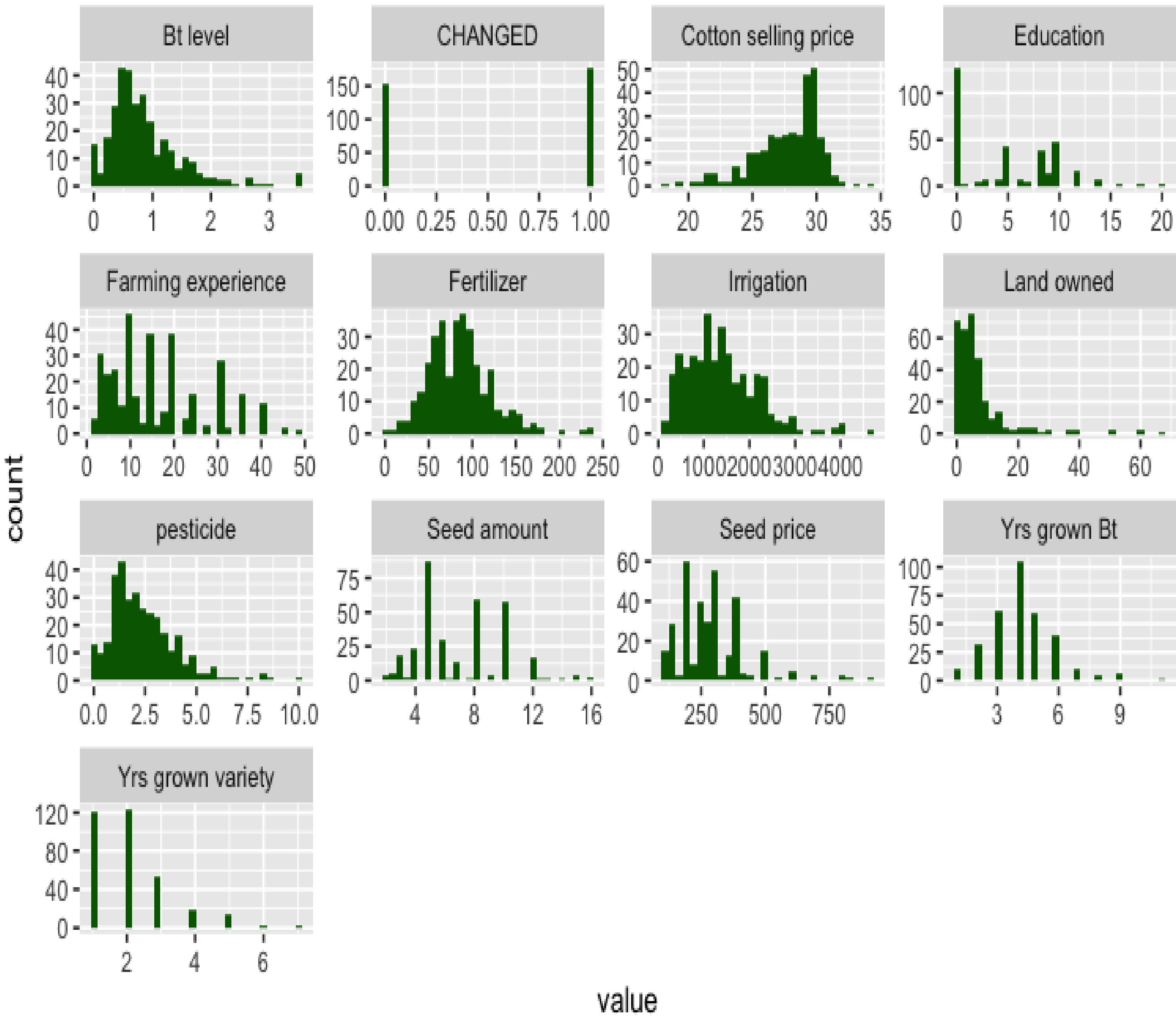

Figure C1: Distribution of Variables

Figure C1 illustrates the distribution of the key variables used in the empirical methodology, across the 331 farmers who are the focus of analysis. Values are on the x-axis while counts are on the y-axis. For example, the first plot shows that Bt content ranges between 0 and 3.5 micrograms of the Bt protein per gram, with the most common value (mode) for a farmer being about 0.5.

Appendix D: Alternative explanations & robustness

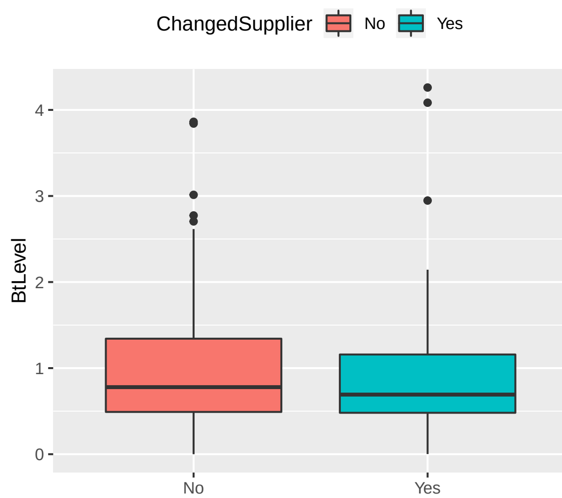

Figure D1: Supplier change

Figure D1 considers behavioral responses to learning besides variety switching and shows it is unlikely that farmers reacted to low Bt level by switching suppliers in the next year.

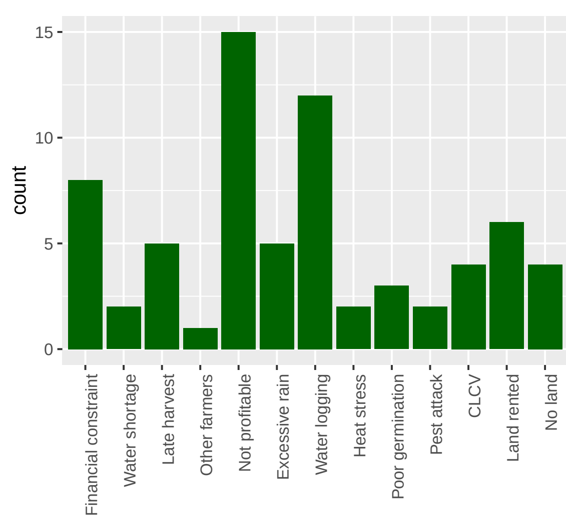

Figure D2: Reasons for exit

Figure D2 rules out the possibility that farmers reacted to low Bt levels by exiting cotton production, by showing the reasons the farmers who exited gave for their decision.

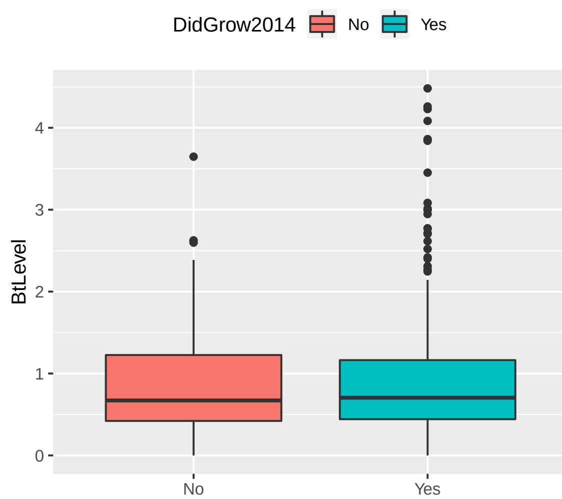

Figure D3: Correlation between Bt level and exit

Figure D3 complements the result in Figure D.2 and shows no correlation between the farmer's Bt level and the decision to exit production.

Table D1: Learning from others

	Dependent variable:			
	CHANGED			
	(1)	(2)	(3)	(4)
Bt level (standardized)	−0.022	−0.029	−0.032	−0.039
	(−0.106, 0.062)	(−0.113, 0.056)	(−0.118, 0.053)	(−0.123, 0.045)
Diff Bt Neighbor (non-st.)	0.129	0.148	0.130	0.221**
	(−0.103, 0.361)	(−0.089, 0.384)	(−0.090, 0.349)	(0.001, 0.442)
Education		0.005	0.016**	0.016**
		(−0.007, 0.018)	(0.003, 0.029)	(0.002, 0.029)
Farming experience			0.008**	0.007**
			(0.002, 0.014)	(0.0002, 0.013)
Years variety grown			−0.093***	−0.101***
			(−0.155, −0.031)	(−0.162, −0.039)
Years Bt grown			−0.056***	−0.052***
			(−0.094, −0.019)	(−0.089, −0.014)
Land owned			−0.010***	−0.010***
			(−0.016, −0.003)	(−0.016, −0.004)
Seed price		−0.0003	−0.0004	−0.0003
		(−0.001, 0.0003)	(−0.001, 0.0001)	(−0.001, 0.0002)
Cotton selling price		−0.023	−0.024	−0.027
		(−0.055, 0.009)	(−0.057, 0.008)	(−0.060, 0.006)
Irrigation				−0.0001*
				(−0.0002, 0.00001)
Fertilizer				−0.0005
				(−0.002, 0.001)
Seed amount				0.038***
				(0.013, 0.064)
Labor				0.0004
				(−0.0003, 0.001)
Pesticide				−0.00001
				(−0.0001, 0.00003)
District FE	Yes	Yes	Yes	Yes
Observations	260	260	260	260
R^2	0.124	0.135	0.227	0.266
Adjusted R^2	0.066	0.067	0.151	0.176
Residual Std. Error	0.478 (df = 243)	0.478 (df = 240)	0.455 (df = 236)	0.449 (df = 231)

Note: *p<0.1; **p<0.05; ***p<0.01

Table D1 includes a rough measure of peer effects: the average Bt of farmers who cultivated a different variety in 2013. With social learning, this coefficient would be positive while own-Bt would have a negative coefficient.

Table D2: Using pseudo-Bt measure for all farmers

	Dependent variable:			
	CHANGED			
	(1)	(2)	(3)	(4)
Pseudo Bt (standardized)	−0.017	−0.016	−0.015	−0.019
	(−0.079, 0.045)	(−0.076, 0.045)	(−0.076, 0.046)	(−0.080, 0.043)
Education		0.010^{*}	0.015^{***}	0.012^{**}
		(−0.001, 0.020)	(0.004, 0.026)	(0.001, 0.023)
Farming experience			0.006^{**}	0.005^{*}
			(0.001, 0.011)	(−0.0005, 0.010)
Yrs grown variety			-0.083^{***}	-0.083^{***}
			(−0.131, −0.035)	(−0.132, −0.035)
Yrs grown Bt			−0.016	−0.018
			(−0.047, 0.016)	(−0.049, 0.013)
Land owned			-0.009^{***}	-0.009^{***}
			(−0.015, −0.003)	(−0.015, −0.003)
Seed price		−0.0002	−0.0004	−0.0003
		(−0.001, 0.0002)	(−0.001, 0.0001)	(−0.001, 0.0002)
Cotton selling price		−0.016	−0.019	−0.020
		(−0.042, 0.009)	(−0.045, 0.007)	(−0.047, 0.006)
Irrigation				-0.0001^{*}
				(−0.0002, 0.00000)
Fertilizer				0.0002^{*}
				(−0.00002, 0.0004)
Seed amount				0.025^{**}
				(0.001, 0.050)
Labor				−0.00001
				(−0.0002, 0.0002)
Pesticide				−0.00001
				(−0.00002, 0.00001)
District FE	Yes	Yes	Yes	Yes
Observations	378	378	378	378
R^2	0.177	0.189	0.242	0.269
Adjusted R^2	0.121	0.127	0.174	0.192
Residual Std. Error	0.467 (df = 353)	0.465 (df = 350)	0.452 (df = 346)	0.448 (df = 341)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table D2 combines single and multiple variety farmers who believed they were purchasing Bt. It replicates the main regression in Equation (5.1) but with Bt level, including for single-variety farmers, constructed as an average of the Bt of all other farmers with that variety. This facilitates comparison with the multiple-variety group.

Table D3: Farmer characteristics - In sample vs out of sample

Statistic	Out of sample, N=396	In sample, N=331	p. overall
Head Age	47.4 (12.1)	46.4 (11.3)	0.250
Head Education	4.37 (4.61)	5.02 (4.75)	0.067
Farming Experience	14.5 (13.1)	15.8 (11.1)	0.150
Main Plot Area	5.73 (6.75)	6.68 (11.0)	0.174
Land Owned	5.78 (10.4)	6.75 (9.27)	0.185
Province:			<0.001
PUNJAB	268 (67.7%)	291 (87.9%)	
SINDH	128 (32.3%)	40 (12.1%)	

Table D3 compares key characteristics of the farmers in the sample, $N = 331$, to all the other farmers that were not included in the sample but were part of the Pakistan Cotton Survey, $N = 396$ (total $N = 727$). For the non-region variables, means are provided with the standard deviation in brackets. The last column reports the p-value for the null hypothesis that the means are the same for both groups.

Table D4: Accounting for measurement error

	Dependent variable:		
	CHANGED		
	OLS	*IV*	*OLS*
	(1)	(2)	(3)
Bt - leafs	-0.006		
	$(-0.068, 0.055)$		
Bt - instrumented		-0.034	
		$(-0.180, 0.112)$	
Bt - correlated			-0.067
			$(-0.203, 0.068)$
Education	0.014^{**}	0.015^{**}	0.023^{*}
	$(0.002, 0.026)$	$(0.003, 0.028)$	$(-0.0003, 0.047)$
Farming Experience	0.007^{**}	0.008^{***}	0.017^{***}
	$(0.001, 0.012)$	$(0.002, 0.014)$	$(0.006, 0.028)$
Yrs grown variety	-0.081^{***}	-0.085^{***}	-0.105^{**}
	$(-0.132, -0.031)$	$(-0.140, -0.031)$	$(-0.200, -0.011)$
Yrs grown Bt	-0.041^{**}	-0.045^{***}	-0.142^{***}
	$(-0.073, -0.009)$	$(-0.079, -0.011)$	$(-0.197, -0.087)$
Land owned	-0.008^{**}	-0.008^{**}	-0.008
	$(-0.014, -0.002)$	$(-0.014, -0.002)$	$(-0.018, 0.002)$
Seed price	-0.001^{*}	-0.001^{**}	-0.001^{**}
	$(-0.001, 0.00002)$	$(-0.001, -0.00002)$	$(-0.002, -0.0001)$
Cotton selling price	-0.022	-0.028^{*}	-0.032
	$(-0.051, 0.006)$	$(-0.058, 0.001)$	$(-0.074, 0.010)$
District FE	Yes	Yes	Yes
Observations	331	331	74
R^2	0.276	0.269	0.634
Adjusted R^2	0.200	0.188	0.431
Residual Std. Error	0.447 (df = 299)	0.449 (df = 299)	0.368 (df = 47)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table D4 demonstrates the results from reconstructing the Bt variable to reduce measurement error and re-estimating the effect of Bt content on variety change. Column 1 reconstructs Bt content as an average, for each farmer, of the leaf values only because they are more strongly correlated with each other than boll values. Column 2 uses one leaf value as an instrument for the other to eliminate (the correlated) measurement error. Column 3 keeps Bt content as the average of the leaf and boll values but applies it only to a limited set of observations where the two leaf values are almost identical.

Table D5: Additional robustness checks

	Dependent variable: CHANGED			
	(1: LPM)	(2: LPM)	(3: LPM)	(4: Logit)
Bt level (standardized)	0.016	0.007	0.016	0.055
	(−0.079, 0.111)	(−0.060, 0.073)	(−0.047, 0.079)	(−0.286, 0.172)
Bt level squared	−0.0004			
	(−0.036, 0.035)			
Education	0.014**	0.013**	0.015**	0.067**
	(0.002, 0.025)	(0.001, 0.025)	(0.003, 0.026)	(0.009, 0.128)
Bt level*Education		0.003		
		(−0.007, 0.013)		
Farming experience	0.006**	0.006**	0.006**	(0.033)**
	(0.001, 0.012)	(0.001, 0.012)	(0.001, 0.011)	(0.007, 0.061)
Years variety grown	−0.082***	−0.082***		−0.404***
	(−0.132, −0.031)	(−0.132, −0.031)		(−0.662, −0.161)
Years Bt grown	−0.040**	−0.040**	−0.041**	−0.202**
	(−0.072, −0.008)	(−0.072, −0.008)	(−0.074, −0.009)	(−0.376, −0.035)
Land owned	−0.008**	−0.008**	−0.008**	−0.037**
	(−0.014, −0.002)	(−0.014, −0.002)	(−0.014, −0.002)	(−0.067, −0.008)
Seed price	−0.001*	−0.001*	−0.001**	−0.003**
	(−0.001, 0.00002)	(−0.001, 0.00001)	(−0.001, −0.0001)	(−0.005, −0.001)
Cotton selling price	−0.022	−0.021	−0.015	−0.135*
	(−0.050, 0.007)	(−0.050, 0.007)	(−0.043, 0.013)	(−0.296, 0.020)
Variety grown dummies	No	No	Yes	No
District FE	Yes	Yes	Yes	Yes
Observations	331	331	331	331
R^2	0.279	0.280	0.300	
Adjusted R^2	0.202	0.203	0.214	
Residual Std. Error	0.446 (df = 298)	0.446 (df = 298)	0.443 (df = 294)	

Note: *p<0.1; **p<0.05; ***p<0.01

Table D5 introduces different specifications to Column 3 in Table 2 to test the effect of Bt content on variety change. Column 1 adds a Bt squared variable to allow for nonlinear effects, Column 2 adds an interaction term between Bt content and education to allow for different effects by education, Column 3 uses a sequence of dummy variables the planting history (omitted from table) to allow for nonlinear effects, and Column 4 uses a bias-reducing logit instead of a linear probability model

.

Table D6: Ordered logit to check effect of Bt on perception formation

	Dependent variable:
	Perception (Ordered)
	Logit
Bt level (standardized)	−0.022
	(−0.302, 0.258)
Education	0.019
	(−0.035, 0.072)
Farming experience	0.008
	(−0.017, 0.033)
Years variety grown	0.080
	(−0.141, 0.302)
Years Bt grown	−0.035
	(−0.131, 0.192)
Seed price	−0.001
	(−0.002, 0.002)
District and sowing-time FE	Yes
Observations	331

*p<0.1; **p<0.05; ***p<0.01

Table D6 estimates the effect of Bt on farmer perceptions by including all three levels of farmer perceptions in the dependent variable, with an ordered logit. This serves as a check on the main results in Table 4, which uses a linear probability model and clusters perceptions into a binary 'poor/moderate' versus 'very good' variable..