Imperfect information and learning: Evidence from cotton cultivation in Pakistan

By Amal Ahmad*

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 and underutilized 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.

JEL: D83, O12, O33

Keywords: economic development, imperfect information, information acquisition, learning, learning by doing, transfer problems

* Department of Economics, University of Massachusetts Amherst, 112A Gordon Hall, 418 N. Pleasant Street, Amherst MA 01002. Email - amalahmad@umass.edu. I would like to thank James Heintz, Daniele Girardi, Emily Wang, Peter Skott, Arslan Razmi, Mariam Majd, and the participants of the Analytical Political Economy workshop at UMass Amherst for very helpful feedback and comments on earlier drafts of this paper. I also thank David Spielman of the International Food Policy Research Institute for graciously answering questions about the dataset. All remaining errors are my own. Declarations of interest: none.

I. Introduction

Economic development is a process characterized by potentially severe and persistent information failures for both private and public agents. Imperfect information can negatively impact the development of the credit market (Weiss and Stiglitz, 1981), the wage labor market (Shapiro and Stiglitz, 1984), the acquisition and adoption of technology (Evenson and Westphal, 1995), and the success of governance strategies (Khan, 2012). In developing countries, these challenges are pronounced, and successful economic development is by definition the process by which these challenges, among others, can be overcome.

Given the salience 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 and environmental conditions, 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). This, 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, 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; 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 I apply it to a rich and underutilized 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 her 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 is the fourth largest grower and exporter of raw cotton in the world, making cotton central to the country's economic development, but farmers have limited information about the pest resistance technology of the varieties they purchase. *Bacillus thuringiensis* (Bt) cotton, which is a genetically modified crop engineered by US-based Monsanto in 1996 to be toxic to bollworm pests, is the most popular type of cotton in Pakistan. However, it was introduced in the country haphazardly, through unlicensed borrowing of the original Monsanto-developed Bt variety and trial-and-error mixing with local varieties (Spielman et al, 2017). Local hybridization was accompanied with poor breeding methods and improper genetic checks, and there are now many "Bt" varieties in Pakistan with different amounts of Bt effectiveness (pest toxicity) (Spielman et al, 2015).

As a result of haphazard technology adoption and the weak capacities of the state, the Pakistani government has failed to ensure that cotton varieties are accurately labeled, standardized, or verifiable in the seed market, and seeds are often sold without appropriate packaging or labelling. Most cotton farmers purchase seeds and rely on the seller to tell them what the variety is and whether it is Bt effective or not, without being able to verify the information. To the extent that Bt effectiveness improves yield outcomes, missing knowledge about this key input is an impediment to productivity.

Using a representative sample, the PCS survey team tested the level of Bt 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 they 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 pest resistance and vice versa.

The results show that farmers are unable to learn about biophysical resistance by observing cultivation outcomes, with Bt content informing neither farmer perceptions of pest resistance nor seed selection next year. Regression analysis also confirms this is because it is difficult for them to distinguish the biophysical characteristics of the plant from environmental conditions that also affect performance.[1] 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%

[1] The data is insufficient for rigorously testing learning from peers, but I use robustness checks in Section 7 to confirm that the results do not change when a rough measure of peer effect is included.

of industry value in 2013-2014. Therefore, this paper has implications not only for microeconomic behavior but also for productivity growth at the macro 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).[2] 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,[3] and heterogeneity in growing conditions can mute social learning or peer effects (Foster and Rosenzweig, 2010).

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 (Nelson, 1970; 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.[4] Since the government can greatly ameliorate the information

[2]These studies measure learning as change in input levels or adoption rates, but a growing literature assesses farmers' discovery of the "price" the good, or the willingness to pay for it, using experimental auctions (Tjernstrom, 2014; Waldman, 2014; Morgan, 2018).

[3]It is particularly appropriate for this study; the farmers indicate in their answers the near absence of any help from NGOs, farmer cooperatives, or other sources of extension services.

[4]Few studies address the credence goods problems in developing countries, and none except Auriol

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, it contributes to the development literature more broadly by demonstrating how information problems generated in the technology acquisition stage can trickle down and hinder effective use after adoption, and how they can compound information failures in other markets. The information problem in this case emerged during the acquisition of the Bt gene, due to constraints on effective local adaptation and governance.[5] Challenges with technology 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 can persist post-acquisition. The paper's results demonstrate high persistence in one such market. In addition, the paper may point to how these information failures can spill into other markets and deepen other information problems. For example, in a rural financial market with principal-agent problems, the agent, if borrowing to purchase inputs, may face difficulty evaluating input quality and the ability to pay back the loan. In this case, the information problem will extend beyond information asymmetry and incentive-compatible mechanisms 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. The next section 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

and Schilizzi (2015) focuses on agricultural seeds in developing countries. Even that paper is a theoretical investigation of how the costs of certification distort equilibrium and not an empirical application.

[5] 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 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.

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, and section 8 summarizes and concludes.

II. 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. By 2019, it is estimated that over 1.6 million farmers cultivate cotton in Pakistan; cotton cultivation accounts 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[6] *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, Pakistani farmers began introgressing this specific gene into local germplasm to create locally specific hybrid

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

Bt varieties.[7] Local Pakistani farmers were therefore able to use trial and error and mixing with local germplasm to "effectively" introduce Bt to their cotton crop, despite intellectual property barriers.

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.[8]

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 in 2013 to what they think they are planting. They find that a

[7] 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.

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

large portion of 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 farmer characteristics and input use are controlled for.

III. Theoretical model

I model how farmers would act once they do or do not learn; in Section 5, I use this to derive an empirical specification to test for learning indirectly by looking at behavioral responses after learning or lack thereof. 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.

Suppose an observable outcome for farmer i at time t, Y_{it},[9] is a function of the unknown level of some attribute B_{it} and of other factors e_{it}: $Y_{it} = f(B_{it}, e_{it})$. 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.[10]

[9] This would be pest damage, for example, in this context.
[10] This deduction process is not necessarily costless, and I discuss this further 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,[11] 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.[12]

I assume there is a surplus in the market each period, with more seeds available for sale than those 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.[13]

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

[11]Throughout, I will refer to B_t as Bt level even though it includes division by the price.

[12]The farmers (from their survey answers) do not store cotton seed for use in the next cultivation period, so that their seed decisions for 2014 are in fact made in 2014. Those who report cultivating the same variety went out and bought that variety again, and are not simply reliant on stored seeds.

[13]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. Therefore, the model can be seen as representing the limiting case of general excess supply, with the "real-world" lying somewhere between the null and this limiting case. The null would be the other limiting case of too-low supply in the second period. Farmers would be constrained from acting on learned information through seed selection, and even if some farmers benefit, average Bt content consumed would be predetermined so there is no improvement in market outcomes.

tion standards; while on average the farmer receives the mean level on sale, so that $E(\tilde{B_{1i}}) = E(B_1)$, 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.[14] 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, $V_1^* = E(B_1) + \gamma$. 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, so that $E(B_2) = E(B_1)$.[15]

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:

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

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 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.[16] Then for those who do not switch,

[14] I address the importance of fixed expectations in the empirics section.

[15] 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 would still hold since $E(\tilde{B_{1i}}) = E(B_1)$.

[16] 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.

their expected second-period Bt content will be

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

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 $V_1^* > B_{1i}$:

$$s_i = \begin{cases} 0 & \text{if } \tilde{B_{1i}} \geq V_1^* \\ 1 & \text{if } \tilde{B_{1i}} < V_1^* \end{cases} \tag{3}$$

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**:

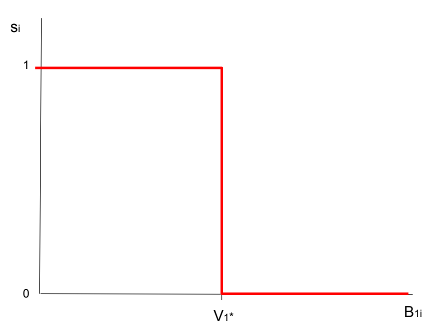


FIGURE 1. DISCRETE SWITCHING

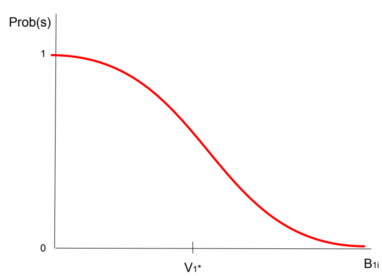
Note: Figure 1 shows whether or not producers switch varieties next year if learning about Bt content is possible. Those who find out Bt content exceeded their expectations do not switch and vice versa.

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).

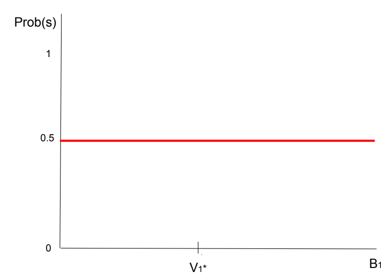
Farmers have numerous competing demands on their attention and need to make many decisions. Moreover, the relative cost versus benefit of exerting attention and processing information to uncover $\tilde{B_{1i}}$ and act accordingly may depend on the absolute difference $|\tilde{B_{1i}}$ - $V_1^*|$. It 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). 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 2a**. 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 $g' < 0$.

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 2b**.



(a) With learning



(b) Without learning

FIGURE 2. PROBABILISTIC SWITCHING

Note: Figure 2 shows switching as a smooth function of Bt content. In Panel a, learning occurs but is easier at the extremes; farmers are more likely to switch the further below expectations Bt content is. In Panel b, it is impossible to deduce Bt and the probability of switching is a tossup for all values.

The expected value of Bt in the next period for each farmer is the expected outcome for switching or not, (1)-(2), weighted by that probability:

$$
\begin{aligned}
E_i(\tilde{B_{2i}}) &= Prob(S)\,[E(B_1)] + (1 - Prob(S))\left[p\tilde{B_{1i}} + (1-p)E(B_1)\right] \\
&= E(B_1)\,[1 - p(1 - Prob(S))] + p(1 - Prob(S))\tilde{B}
\end{aligned} \tag{4}
$$

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 (4):

$$
\begin{aligned}
E_i(\Delta\tilde{B}_i) &= E_i(\tilde{B_{2i}}) - \tilde{B_{1i}} \\
&= E(B_1)\left[1 - p(1 - Prob(S))\right] + p(1 - Prob(S))\tilde{B_{1i}} - \tilde{B_{1i}} \\
&= \left(E(B_1) - \tilde{B_{1i}}\right)\left[1 - p(1 - g(\tilde{B_{1i}} - V_1^*))\right]
\end{aligned}
\tag{5}
$$

Expected change in Bt content across the market is found by taking expected change for each initial realization $\tilde{B_{1i}}$, (5), weighing it by the probability of its occurrence in the first period $Prob(\tilde{B_{1i}})$, and summing across:

$$
\begin{aligned}
E(\Delta\tilde{B}) &= \sum_i Prob(\tilde{B_{1i}}) * E_i(\Delta\tilde{B}_i) \\
&= \sum_i Prob(\tilde{B_{1i}}) * \left(E(B_1) - \tilde{B_{1i}}\right)\left[1 - p(1 - g(\tilde{B_{1i}} - V_1^*))\right]
\end{aligned}
\tag{6}
$$

A simpler expression can be obtained for (6) - 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 $\alpha > 0$;[17] the parameter α is the *learning parameter*. This form guarantees that for values at the expectation switching is a tossup, $g(0) = 0.5$. Furthermore, note that, allowing for a discrepancy between expectations and true market averages, $V_1^* = E(B_1) + \gamma$. Substituting into (6):

(7)

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

(7) can be written in continuous form. Replacing the summation with integration,

[17]This function would be bounded between 0 and 1 to ensure that $0 \leq g \leq 1$.

and letting $f(\tilde{B_{1i}})$ be the probability density function, then:

$$(8) \quad E(\Delta\tilde{B}) = \int f(\tilde{B_{1i}}) \left(E(B_1) - \tilde{B_{1i}}\right) \left[1 - 0.5p + \alpha p(E(B_1) + \gamma - \tilde{B_{1i}})\right] d\tilde{B_{1i}}$$

Simplifying (8) we get the following (detailed calculation in Appendix 1):

$$(9) \quad \begin{aligned} E(\Delta\tilde{B}) &= \int f(B)\,(E(B) - B)\,[1 - 0.5p + ap(E(B) + \gamma - B)]\,dB \\ &= \alpha p\left[E(B^2) - (E(B))^2\right] \\ &= \alpha p Var(B) \geq 0 \end{aligned}$$

Therefore, by assuming a linear form for g, and since $Var(\tilde{B_1}) = \sigma^2$, we get:

$$(10) \qquad E(\Delta\tilde{B}) = \alpha p \sigma^2$$

The probability of switching may also depend on the *standardized* deviation from expected content, so that $g(x) = -\alpha\left[\frac{\tilde{B_{1i}} - V_1^*}{\sigma}\right] + 0.5$. In that case, it is easy to show that (10) is amended as follows:[18]

$$(11) \qquad E(\Delta\tilde{B}) = \alpha p \sigma$$

The model shows that as α - the extent of learning and response - and p - the extent of variety integrity - increase, average Bt quality consumed rises; 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. With

[18] I checked the results from (10)-(11) by simulating learning and probabilistic switching from the sample in my data. For different α, p, and σ, the average change in market outcome $E(\Delta\tilde{B})$ is identical to that predicted in (10)-(11).

additional calculations using the effect of Bt on yield, we can also estimate the extent of monetary benefit to farmers from said improvements in average Bt.

To illustrate, **Figure 3** uses (11) to show how different values of α and p translate to expected Bt improvement $E(\Delta\tilde{B})$, for a fixed $\sigma = 0.57$ (the standard deviation of Bt in my data). For the axes, p ranges from 0 to 1, while α ranges from 0 to 0.20[19] As shown in the figure, at $\sigma = 0.57$ and an upper limit of $p = 1$ and $\alpha = 0.20$, switching would result in average Bt content improving by 0.12 $\mu g/g$, or about 14% improvement over the mean of 0.88 $\mu g/g$. Appendix 2 shows how, using evidence-based benchmarks on the effect of Bt on yield and revenue, this would translate into gains of roughly about 45 million USD for Pakistani cotton farmers in one year.

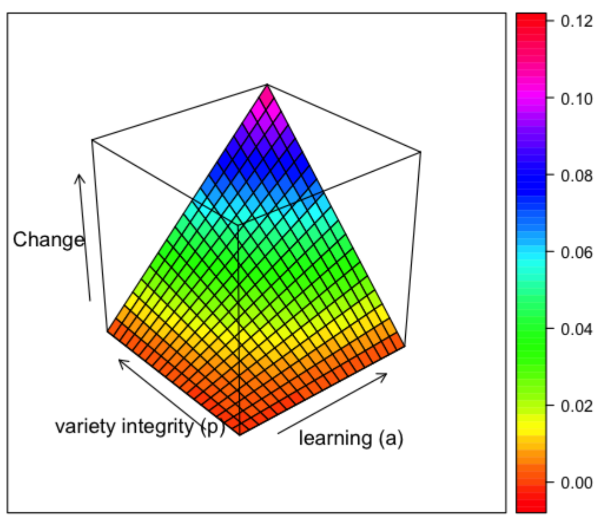


Figure 3. Improvement in Bt averaged across farmers

Note: Figure 3 plots improvements in Bt quality as a function of both the learning parameter a, for which the axis in the figure ranges from 0 to 0.20, and the variety integrity in the market, for which the axis ranges from 0 to 1. On the vertical axis is the change in Bt, measured in $\mu g/g$. It shows that greater learning and variety integrity improve market outcomes.

[19] I am assuming that a plausible upper limit is that farmers who find their Bt content is one standard deviation above what they expected are 20% less likely to switch varieties.

In Section 5 I use this model 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 empirical tests show that α is positive (Figure 2a), we can conclude that farmers learn from experience and respond accordingly. However, while absence of learning necessarily generates a zero slope (Figure 2b), 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. 1) and from not switching (Eq. 2) become equivalent, and the profit maximization exercise that results in switching is not valid.[20] [21]

Therefore, the empirical exercise 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. A positive α should be accompanied by analysis showing that Bt content systematically influences perceptions of the variety's resistance and the opposite for null α.

[20] On a macro level, note that regardless of the value of α, if $p = 0$ average Bt content will remain the same and there will be zero benefit to farmers from any switching strategy.

[21] Farmers would also not act if they believe there is very little variance in the quality on the market, so if they believe $\sigma^2 \approx 0$. However, I am ruling this out because with fixed farmer expectations, believing there is zero variance would imply $B_{1i} = B_1 = V_1^*$ for all farmers, i.e. zero actual variance.

IV. 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) in conjunction with local agricultural scientists between March 2013 and January 2015, on a random stratified sample of 727 farmers in the provinces of Punjab and Sindh. These provinces account for 99% of all cotton production in the country, and the sample is nationally representative. The surveys are accessible publicly from the Harvard Dataverse website.

The first survey, Round 1.1, collected preliminary background data on 727 cotton 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 sowing was complete for the season but before harvest was completed. Only 601 of the farmers ended up sowing cotton. 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.[22]

The third survey, Round 1.3, followed up in January 2014 and at this time the last picking was complete. The farmers were asked about input use, quantities harvested and sold, revenue, and performance perceptions, as well as assets owned, general consumption patterns, and decision-making by gender.[23]

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 535 of those who

[22]Farmer answers in this round show that social networks such as farmer coops and NGOs are nearly nonexistent and that the use of cash credit is also negligible.

[23]Nearly all household heads and decision makers in this dataset are male and the women are not involved in the decision making process and, in most households, in the cotton labor process either.

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 where they were tested for the presence of specific genes and toxins that contribute to Bt expression; the exact methodology is detailed in Ma et al (2017). The farmers were not made aware of the results until early 2015, by which point the 2014 growing season was 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).

V. Econometric methodology

To measure the extent of learning α, I regress variety switching in the next year on standardized Bt content, or $\frac{\tilde{B_{1i}}-E(B_1)}{\sigma}$, in the current year. This generates a regression coefficient which is the slope of the function in Figure 2.[24] Bt content is seed-price adjusted by including seed price as a control. 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.[25] The main regression is:

$$Change_i = \beta_0 + \beta_1 BtLevel_i + \sum \beta_j Controls_{ji} + \epsilon_i \tag{12}$$

Change takes a value of 1 if the farmer switched varieties in 2014 and 0 otherwise, *Bt level* is the (standardized) Bt effectiveness of the farmer's 2013 variety

[24] Technically, it assumes that the function is linear, so that the slope is constant.

[25] It is for this reason that I need a homogenous V_1^*. Otherwise the regression is undoable without farmer-specific fixed effects, untenable because this is not panel data. To ensure this assumption is plausible also informed sample selection, I only run the regression on farmers who when they bought the seed said they think it's a Bt seed. Those who said they do not think it is, or don't know, are excluded.

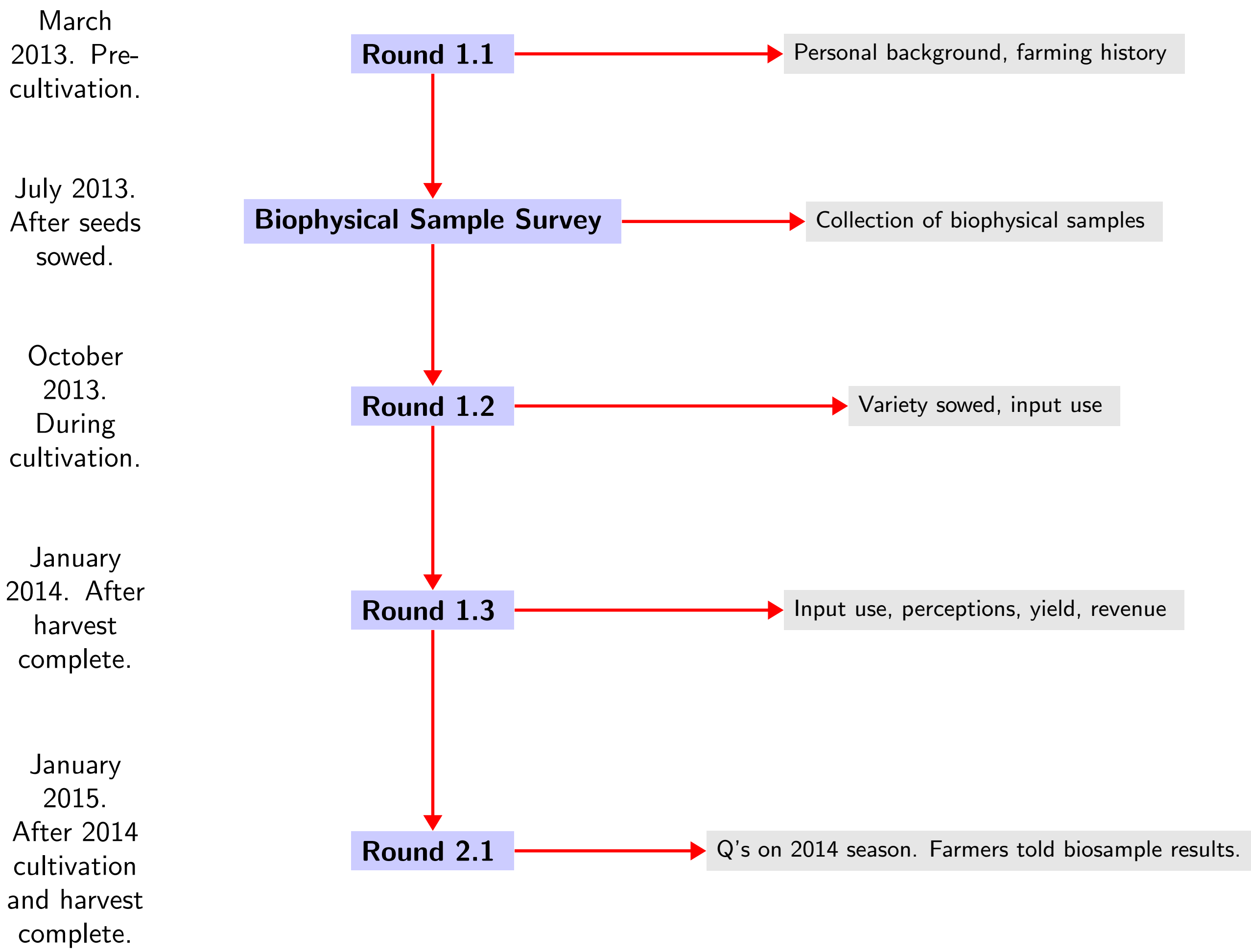


Figure 4. Timeline of surveys and cultivation

Note: 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.

as measured by the Biophysical Sample Survey, and controls are other factors, occurring in 2013 or beforehand, 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 $\beta_1 = 0$.

Given acceptable controls, identification is straightforward. Bt level is measured for the 2013 variety while the choice to change varieties is made the following year. Even without the time lapse, feedback in the other direction is ruled out: 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.

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 such that picking the same variety again is likely to result in a similar Bt level again. 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 $p > 0$, in the theoretical model), we would expect them to switch varieties in 2014 if they felt their 2013 variety had poor resistance, all else constant. The relevant regression is:

$$Change_i = \gamma_0 + \gamma_1 ResistancePerception_i + \sum \gamma_j Controls_{ji} + \epsilon_i \tag{13}$$

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 supports the behavior outlined in the theoretical model, so that a null β_1 in Equation (12) signals 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:

$$(14) \qquad ResistancePerception_i = \theta_0 + \theta_1 BtContent_i + \sum \theta_j Controls_{ji} + \epsilon_i$$

$\theta_1 > 0$ 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 (14) to be associated with $\beta_1 < 0$ in Equation (12). 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 then imply no impact of Bt on variety choice ($\beta_1 = 0$).[26]

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:

$$(15) \qquad Pesticide_i = \phi_0 + \phi_1 Btcontent_i + \sum \phi_j Controls_{ji} + \epsilon_j$$

Pesticide measures effective pesticide use per acre in 2013. 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. However, this regression is only valid if farmers can learn about Bt content *before* cultivation is over, so that there is room for adjusting input decisions in the same season. This is not highly plausible, as it is more likely that input use in that season is predetermined relative to Bt content. Therefore Equation (15) is not the focus of the discussion but used as a supplemental result. **Table 1** summarizes the possible coefficient combinations and interpretations.

[26] In both scenarios, I expect $\gamma < 0$, to signify that farmers do believe there is some market integrity and would act as the model predicts.

Table 1—Coefficient combinations and interpretation

β_1 : Bt on variety Δ	γ_1 : Perception on variety Δ	θ_1 : Bt on perception	ϕ_1 : Bt on pesticide use	**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 use.
$=0$	<0	$=0$	$=0$	Farmers are unable to learn.

Note: Table 1 summarizes the possible meaningful combinations of coefficients and their corresponding economic interpretations. The coefficients in Columns 1-4 are derived from Equations (12)-(15).

Key variables are constructed as follows. *Change* is binary: it takes a value of 1 if the farmer changed the variety they cultivated between 2013 and 2014, and 0 if they cultivated the same variety both years. *Resistance perception* is a three-level dummy variable connoting "poor", "moderate", and "very good". When used as a dependent variable in Equation (14), however, I cluster it into binary levels "poor/moderate" and "very good" to allow for a linear probability model; I check that results are similar with an ordered multi-level logit. *Pesticide* measures effective pesticide use for the 2013 season per acre, which I construct by weighing the quantity of different pesticides used (in mL) by their percentage strengths, adding them, and dividing by acres of cotton cultivated.

The main independent variable of interest, *Bt level*, is the standardized expression level of the Bt *cry* protein as measured by the labs' sandwich ELISA tests for the 2013 varieties. It is measured in micrograms of the protein per gram of leaf tissue. A higher level indicates more toxin, therefore a greater effectiveness in targeting and eliminating bollworms. 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. My variable is an average of the measurements 70 days after sowing for each variety.[27]

For controls, in Equations (12) and (13) I control for other factors that can affect variety selection: (i) farmer characteristics that may influence how the farmer deals with his crop (education, years of general farming experience, years of experience cultivating what he thinks is Bt cotton, land owned by 2012 as proxy for wealth); (ii) planting history (the number of years the 2013 variety has been cultivated); (iii) price per unit of the seed variety purchased for the 2013 season; (iv) price per unit of post-harvest cotton fetched by the 2013 variety;[28] (v) input intensity (irrigation, Nitrogen fertilizer, labor, and seeds sowed, all per acre of cotton cultivated); (vi) 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*.

For Equation (14), environmental factors that affect resistance performance such as exogenous pest intensity need to be controlled for but information for this is not available. Since it appears that pest intensity is time- and space-dependent, I assume that controlling for the time-of-sowing[29] and the geographical district can roughly control for general exogenous pest intensity. I include these controls as well as personal characteristics, planting history, seed price, and input intensity. For Equation (15), I control for farmer characteristics and planting history, as well as the area cultivated, seeds per acre sowed (more concentrated cultivation per acre requires more pesticide per acre), soil type (since it can impact pesticide absorption), and sowing time and district to control for exogenous pest intensity. Appendix 3 provides further details on how the control variables are constructed.

Table 2 provides a summary of the distribution of key variables in the data.

[27]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.

[28]This is distinct from yield performance; it captures desirable qualities in the variety such as whiteness of the cotton, quality of the lint, and length of the staple fiber.

[29]I have data on time of sowing by 10 day periods, for example early March, mid March, etc.

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. With regard to perception, 18% of farmers rated the bollworm resistance performance as poor, 40% as moderate, and 42% as very good. The histograms in **Figure 5** illustrate these and other distributions.

Table 2—Distribution of Variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Changed	301	0.558	0.497	0	0	1	1
Bt level (non-standardized)	301	0.877	0.692	0.000	0.484	1.163	4.260
Education (years)	301	4.977	4.811	0	0	9	20
Farming experience (years)	301	16.213	10.768	2	8	22	50
Years variety grown	301	2.023	1.091	1	1	2	7
Years Bt grown	301	4.110	1.735	1	3	5	11
Land owned (acres)	301	6.516	8.859	0.00	1.50	7.00	67.00
Seed purchase price (PR per kg)	301	280.7	135.1	83.3	200	350	840
Selling price (100 PR per 40 kg)	301	27.863	2.542	18	26.4	29.7	34
Irrigation (mins per acre)	301	1,385	771	120	810	1,840	4,620
Fertilizer (kg per acre)	301	85.598	36.159	0	59.7	103	236
Seed amount (kg per acre)	301	7.01	2.81	2.00	5.00	9.00	16.00
Labor (hours per acre)	301	161.3	90.3	36.0	104.0	197.3	967.5
Pesticide (effective L per acre)	301	0.563	0.430	0	0.285	0.725	2.437

Note: Table 2 summarizes the distribution of the key variables used in the analysis.

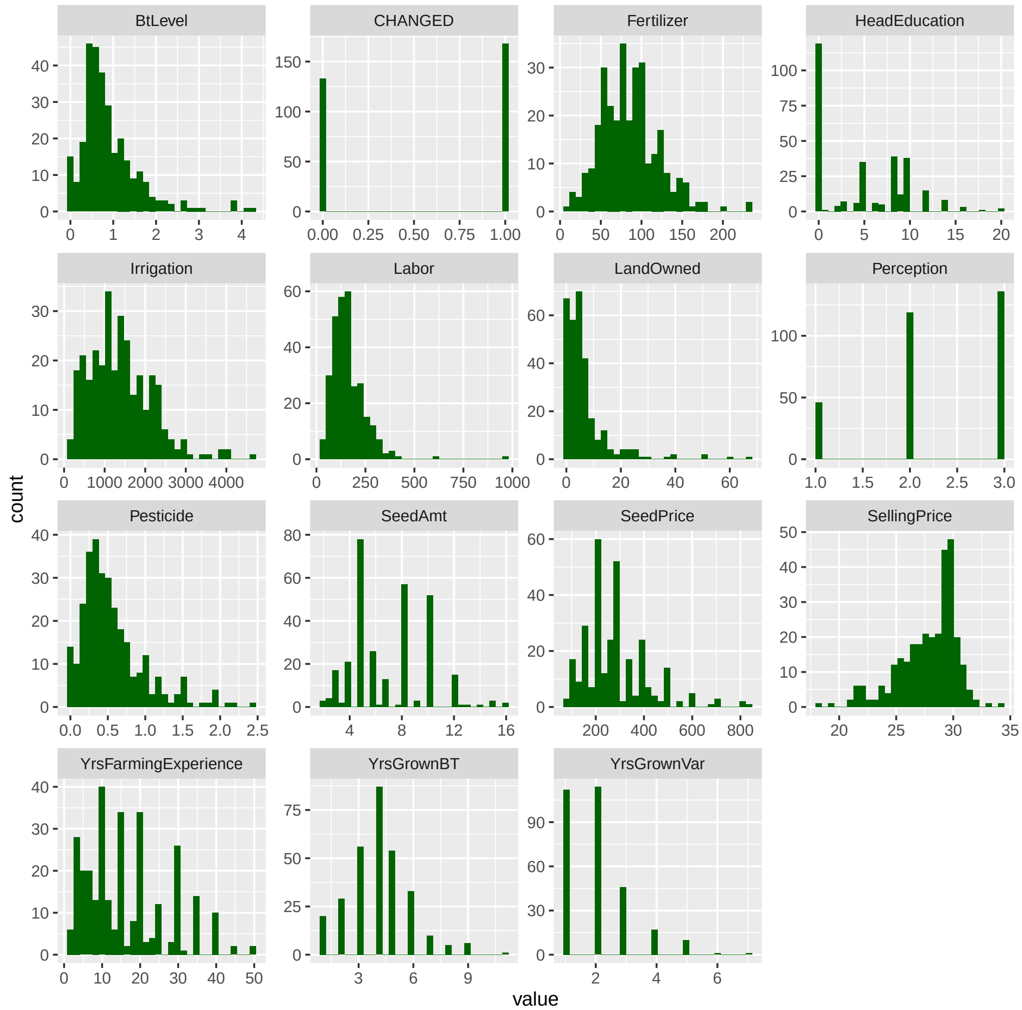


Figure 5. Distribution of key variables

Note: Figure 5 illustrates the distribution of the key variables used in the empirical methodology, across the 301 farmers in the sample. 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 4 micrograms of the Bt protein per gram, with the most common value (mode) for a farmer being about 0.5.

The above empirical exercises necessitate that I restrict my sample to 301 households. First, only 535 households sowed cotton in both 2013 and 2014.[30] Out of those, I need households that farmed only one cotton variety on the plot from

[30]Those that did not cited reasons environmental reasons such as high water logging or personal reasons such as preference for another crop; none cited problems with bollworms.

which the biophysical sample was taken in 2013, because for farmers that cultivated more than one variety it is impossible to tell which variety the lab tests correspond to. In addition, I restrict my sample to those where the biophysical sample was taken from the "main" plot, as each farmer can farm multiple plots but the data is more comprehensive about what each farmer identifies as their main plot. Finally, it is important to fix ex-ante expectations, in line with the theoretical model. I focus only on farmers who believed they were cultivating a Bt variety in 2013, answering 'Yes' when asked if their variety is Bt effective.[31] These considerations restrict my sample to 301 households.

Table 3 shows that the farmers in sample (301) and out (426) are similar average age, education, years of farming experience, the area of the main plot they operate on, and the total area of land they own. The significant difference is province: the in-sample group is more heavily skewed toward Punjab than Sindh. Therefore, the results from the in-sample regressions are likely representative, at least roughly, of farmers in the survey, who are in turn nationally representative.

VI. Results and discussion

Diagrams showing the relationship between key variables can provide a sense of correlation, paving the way for empirical analysis. **Figure 6a** shows the absence of correlation between Bt level and variety change, while **Figure 6b** shows the absence of correlation between Bt level and resistance performance perception. In both, an ANOVA test fails to reject the null hypothesis of no difference in mean Bt level amid different groups. By contrast, **Figure 6c** documents correlation between resistance perception and variety change: farmers who perceive resistance performance as better are less likely to change variety in the next year. Finally, **Figure 6d** shows no correlation between Bt level and pesticide use.

These correlations, if causal, support the hypothesis in line 4 of Table 1: that

[31] Intuitively, including farmers who did not think they were purchasing Bt would invalidate the exercise, since they would all be assuming $V_1^* = 0$ so that no switching would be predicted regardless of learning.

TABLE 3—FARMER CHARACTERISTICS - IN SAMPLE VS OUT OF SAMPLE

Statistic	Out of sample, N=426	In sample, N=301	p. overall
Province:			<0.001
PUNJAB	297 (69.7%)	262 (87.0%)	
SINDH	129 (30.3%)	39 (13.0%)	
Head Age	47.3 (12.0)	46.3 (11.4)	0.257
Head Education	4.45 (4.59)	4.98 (4.81)	0.136
Farming Experience	14.6 (12.9)	15.9 (11.1)	0.134
Main Plot Area	6.18 (9.82)	6.14 (7.59)	0.944
Land Owned	6.01 (10.6)	6.52 (8.86)	0.483

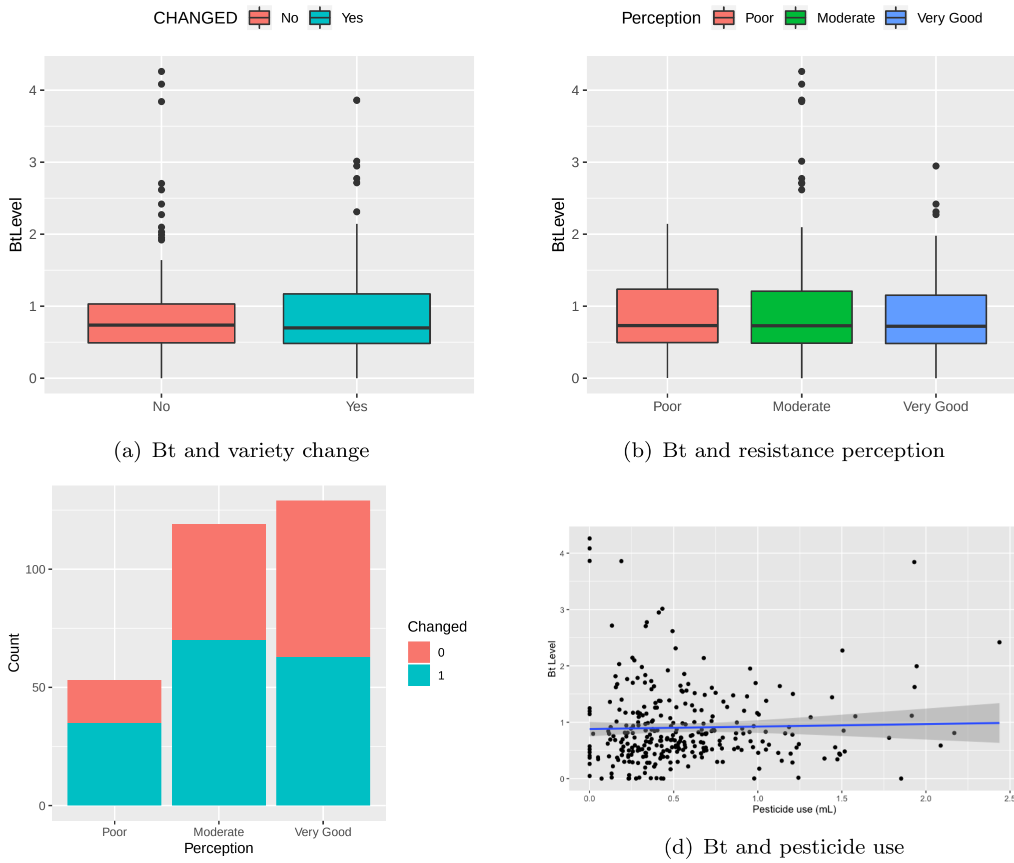
Note: Table 3 compares key characteristics of the farmers in the sample, $N = 301$, to all the other farmers that were not included in the sample but were part of the Pakistan Cotton Survey, $N = 426$ (total $N = 727$). For the non-region variables, means are provided with the standard deviation in brackets. The last column reports the p-value for the null hypothesis that the means are the same for both groups. A sufficiently small p-value implies rejection of this null.

farmers believe switching is an effective strategy to improve resistance ($\gamma_1 < 0$) but are unable to learn the relevant biophysical attribute from cultivation experience ($\theta_1 = 0$) and therefore to incorporate it into variety selection ($\beta_1 = 0$) or input use ($\phi_1 = 0$). Next, I show that the empirical results arrive at the same conclusions.

EFFECT OF BT ON VARIETY CHOICE. — **Table 4** shows the results from four versions of Equation (12). All are linear probability models to facilitate interpretation, with robust standard errors (adjusted for heteroskedasticity). The 95% confidence intervals are noted below each coefficient.

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,[32] seed purchase price, and cotton selling price. As argued in the theoretical model, controlling for seed price is important and Bt content should be price-adjusted. Meanwhile, cotton

[32]Ma et al (2017) show that education is the only farmer characteristic correlated with Bt levels in this sample; specifically, more educated farmers are likely to have seeds which have higher Bt expression, a difference that is small in magnitude but significant statistically.



(a) Bt and variety change

(b) Bt and resistance perception

(c) Resistance perception and variety change

(d) Bt and pesticide use

Figure 6. Correlations between key variables

Note: Figure 6 illustrate the correlations between the key independent and dependent variables in Equations (12)-(15). It shows no correlation between Bt level and the following: whether farmers change variety next year (Panel a), how they perceive bollworm resistance performance (Panel b), and how much pesticide they use (Panel d). Panel c shows that improved farmer perceptions are correlated with lower likelihood of variety change the next year.

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 selling price variable.[33] Column 3 adds variables which are exogenous to Bt level but may affect variety choice, whose inclusion would therefore improve precision: farmer characteristics such as farming experience, planting history, and wealth. Column

[33]Cotton selling price is not endogenous to each farmer's production outcomes since the farmers are small and therefore price takers.

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.[34]

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 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 -0.5% to +0.2%, depending on the specification. A 95% confidence interval can rule out negative effects larger than 7% in absolute value across all specifications.

The coefficients on the control variables possess the expected signs. Higher cotton selling prices reduce the chance that the farmer will change the variety the next year, but seed purchase prices have coefficients that are close to zero and insignificant, 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 and insignificant, 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.[35] 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 have an awareness of how to cultivate it more efficiently, reducing the need for regular variety change.

[34] 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.

[35] The one exception is the amount of seeds sowed which is positive and significant. Possibly, varieties sowed more intensely were ones that were failing to grow properly, hence a higher likelihood that the variety will be changed the next year.

Table 4—Effect of Bt level on variety change

	Dependent variable:			
	Changed variety			
	(1)	(2)	(3)	(4)
Bt level (standardized)	0.001	−0.002	0.002	0.004
	(−0.061, 0.064)	(−0.065, 0.062)	(−0.061, 0.064)	(−0.060, 0.068)
Education		0.008	0.014^{**}	0.013^{**}
		(−0.004, 0.020)	(0.003, 0.026)	(0.001, 0.025)
Farming experience			0.008^{***}	0.007^{**}
			(0.002, 0.014)	(0.001, 0.013)
Years variety grown			-0.074^{***}	-0.078^{***}
			(−0.128, −0.020)	(−0.132, −0.023)
Years Bt grown			-0.040^{**}	-0.036^{**}
			(−0.075, −0.006)	(−0.071, −0.001)
Land owned			-0.007^{*}	-0.007^{*}
			(−0.014, 0.001)	(−0.014, 0.0001)
Purchase price (seed)		−0.0002	−0.0004	−0.0003
		(−0.001, 0.0003)	(−0.001, 0.0002)	(−0.001, 0.0002)
Selling price (cotton)		−0.024	−0.023	−0.023
		(−0.054, 0.006)	(−0.054, 0.007)	(−0.055, 0.009)
Irrigation				−0.0001
				(−0.0002, 0.00003)
Fertilizer				−0.0003
				(−0.002, 0.001)
Seed amount				0.030^{**}
				(0.005, 0.055)
Labor				0.0004
				(−0.0002, 0.001)
Pesticide				0.00000
				(−0.00004, 0.00005)
District fixed effects	Yes	Yes	Yes	Yes
Observations	301	301	301	301
R^2	0.217	0.229	0.283	0.307
Adjusted R^2	0.149	0.153	0.201	0.212
Residual Std. Error	0.459 (df = 276)	0.458 (df = 273)	0.445 (df = 269)	0.442 (df = 264)

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Table 4 demonstrates in detail the results of Equation (12), taking into account different control specifications to show robustness. Across specifications, Bt content does not influence variety change in the next year.

Effect of perceptions on variety choice. — **Table 5** shows that the results change significantly when we assess the impact of farmer *perceptions* of bollworm resistance on variety change.[36] 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 improve precision, and Column 3 adds two variables, pesticide use and yield, which may affect *Change* but whose exogeneity to perceptions is not clear.[37]

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 and vice versa. Depending on the specification, farmers who viewed resistance as moderate are 11.2 to 15.2% less likely to change variety in the next year than those who viewed it as poor, and this is significant or almost significant at the 10% level. Farmers who viewed resistance performance as very good are 16.4 to 19.1% less likely to change variety next year than those who viewed it as poor, and this is consistently significant at the 5 or 10% 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 4.

Effect of Bt on perception formation. — **Table 6** explores the role of Bt content in informing farmer perceptions. From the above results, 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.

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,[38] and the key independent variable is standardized Bt content.

[36] Again, the standard errors are robust and district controls are omitted from the presentation.

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

[38] I lump it so I can perform a linear probability model, to maintain ease of interpretation, but I perform robustness checks to ensure consistency of the results with a ordered logit (see Section 7)

Table 5—Effect of perceptions on variety change

	Dependent variable:		
	Changed variety		
	(1)	(2)	(3)
Moderate	−0.112	−0.152*	−0.125
	(−0.271, 0.047)	(−0.307, 0.003)	(−0.287, 0.036)
Very Good	−0.175**	−0.191**	−0.164*
	(−0.346, −0.004)	(−0.354, −0.028)	(−0.332, 0.005)
Education		0.016***	0.016***
		(0.004, 0.028)	(0.004, 0.027)
Farming experience		0.008***	0.008***
		(0.003, 0.014)	(0.003, 0.014)
Years variety grown		−0.072***	−0.074***
		(−0.126, −0.019)	(−0.127, −0.020)
Years Bt grown		−0.044**	−0.045**
		(−0.078, −0.010)	(−0.079, −0.011)
Land owned		−0.007*	−0.007*
		(−0.014, 0.001)	(−0.014, 0.001)
Purchase price (seed)		−0.0004	−0.0004
		(−0.001, 0.0001)	(−0.001, 0.0002)
Selling price (cotton)		−0.021	−0.020
		(−0.052, 0.011)	(−0.052, 0.012)
Pesticide			0.016
			(−0.129, 0.160)
Yield (log)			−0.071
			(−0.171, 0.030)
District fixed effects	Yes	Yes	Yes
Observations	301	301	301
R^2	0.229	0.297	0.302
Adjusted R^2	0.159	0.214	0.212
Residual Std. Error	0.456 (df = 275)	0.441 (df = 268)	0.441 (df = 266)

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

Note: Table 5 demonstrates in detail the results of Equation (13), taking into account different control specifications to show robustness. Across specifications, improved perceptions of bollworm resistance decrease the probability that the farmer will change varieties next year.

Identification comes from the fact that the biophysical attributes can impact farmer perceptions six months later but not vice versa, and multiple specifications are estimated. 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.

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 3% 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 from the table) 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 5 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[39] 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 an 11% greater likelihood of viewing resistance as very good instead of poor/moderate. The results from this column are not causally identified: it is not clear 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.

[39]The effect of Bt content on perceptions would presumably operate through yield.

TABLE 6—EFFECT OF BT LEVEL ON PERCEPTION FORMATION

	Dependent variable:				
	Perception (Very Good)				
	(1)	(2)	(3)	(4)	(5)
Bt level (standardized)	−0.019	−0.021	−0.021	−0.022	
	(−0.071, 0.032)	(−0.073, 0.031)	(−0.074, 0.032)	(−0.076, 0.031)	
Education		0.006	0.006	0.007	
		(−0.007, 0.018)	(−0.007, 0.019)	(−0.006, 0.020)	
Farming experience			−0.0002	0.0002	
			(−0.006, 0.006)	(−0.006, 0.006)	
Years variety grown			0.002	0.001	
			(−0.054, 0.057)	(−0.054, 0.057)	
Years Bt grown			0.003	0.001	
			(−0.033, 0.040)	(−0.036, 0.038)	
Seed price		0.00005	0.0001	0.00005	
		(−0.0004, 0.001)	(−0.0004, 0.001)	(−0.0004, 0.001)	
Pesticide				−0.00002	
				(−0.0001, 0.00003)	
Yield per acre(log)					0.111**
					(0.013, 0.209)
District & sowing-time FE	Yes	Yes	Yes	Yes	Yes
Observations	301	301	301	301	301
R^2	0.284	0.287	0.289	0.292	0.294
Adjusted R^2	0.190	0.187	0.177	0.177	0.201
Residual Std. Error	0.446 (df = 265)	0.447 (df = 263)	0.450 (df = 259)	0.450 (df = 258)	0.443 (df = 265)

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

Note: Table 6 demonstrates in detail the results of Equation (14), taking into account different control specifications to show robustness. In Columns 1-4, Bt content does not systematically influence farmer perceptions of the bollworm-resistance performance of their crop. Column 5 shows that yield is positively correlated with farmer perceptions.

EFFECT OF BT CONTENT ON PESTICIDE USE. — **Table 7** examines the possibility that Bt content can be uncovered and impact not variety choice in the next year but pesticide use in the same season. Identification comes from the fact that the biophysical property of the variety purchased (Bt content) is not driven by the farmer's (subsequent) input choices.

Column 1 regresses effective 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.

Across specifications, Bt content does not affect 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.

Table 7—Effect of Bt level on pesticide use

	Dependent variable:			
	Pesticide use			
	(1)	(2)	(3)	(4)
Bt level (standardized)	−0.007	−0.008	0.001	0.001
	(−0.076, 0.063)	(−0.077, 0.061)	(−0.069, 0.071)	(−0.065, 0.068)
Education		0.012**	0.012*	0.011*
		(0.001, 0.024)	(−0.0002, 0.025)	(−0.002, 0.024)
Farming experience			0.0005	0.001
			(−0.005, 0.006)	(−0.004, 0.006)
Years variety grown			−0.018	−0.018
			(−0.062, 0.025)	(−0.062, 0.026)
Years Bt grown			−0.022	−0.022
			(−0.056, 0.012)	(−0.055, 0.011)
Land owned			0.0002	0.001
			(−0.006, 0.006)	(−0.005, 0.007)
Area Cultivated			−0.001	−0.001
			(−0.006, 0.004)	(−0.007, 0.004)
Irrigation				−0.00001
				(−0.0001, 0.0001)
Fertilizer				0.002**
				(0.0001, 0.003)
Seed amount			0.035**	0.036***
			(0.007, 0.063)	(0.009, 0.063)
District, sowing-time, and soil FE	Yes	Yes	Yes	Yes
Observations	301	301	301	301
R^2	0.303	0.319	0.350	0.366
Adjusted R^2	0.199	0.214	0.232	0.245
Residual Std. Error	0.385 (df = 261)	0.381 (df = 260)	0.377 (df = 254)	0.373 (df = 252)

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

Note: This table demonstrates in detail the results of Equation (15), taking into account different control specifications to show robustness. Across specifications, Bt content does not affect pesticide use.

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 6) and to switch varieties next year accordingly (Table 4). 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 5). Inability to discover Bt content through cultivation may also be evident in the absence of an appropriate response through input use (Table 7).

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 2. Those estimates suggest that if average Bt improves in the long run from the in-sample level of 0.88 $\mu g/g$ to the maximum-effectiveness level of 1.59 $\mu g/g$, 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, whiteness of the fiber is the outcome primarily of the biophysical property of the seed and 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 learn about from cultivation experience alone and in the absence of certification standards or external information provision.

VII. Alternative explanations and robustness checks

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 in the main regressions for two reasons. First, and most importantly, data limitations imply that for each farmer there are very few other farmers within the same village[40] who can therefore be possible peers. Also, there is no information about how far the farmers are from each other, 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 therefore without introducing bias in the regression. Second, in the case of learning about Bt, it is likely difficult enough for a farmer to learn from own experience such that learning from the experience of others is even less likely. Word of mouth from other farmers about what they found may affect a farmer's perceptions, but this too would be based in what the neighbors learned themselves about their own harvest, as opposed to direct observation of peers' choices and outcomes.

Nonetheless, given the role of social learning in many contexts, it is instructive to include a robustness check that incorporates some measure of peer effects. For each farmer, I identify the other farmers in the same village as potential peers. If there is social learning, we would expect a farmer to be less likely to switch the higher is the Bt of peers who purchased the same variety in 2013, and more likely to switch the higher is the Bt of peers who purchased a different variety in 2013. Unfortunately, there is very little information on the former; for most farmers, there are 0 or 1 'peers' in the sample who used the same variety that year. Therefore, I focus on the latter; for each farmer, I calculate the average Bt of other villagers who purchased a different variety in 2013, and expect to see a positive coefficient on switching for this variable if there is social learning. **Table A1** in the Appendix shows the results when this variable is integrated into the main regression of Equation (12).[41] The coefficient on the peer variable is null,

[40] A village is a smaller geographical unit than a district.

[41] The sample size decreases as there are some farmers who do not have other villagers that cultivate a different variety.

as is the coefficient on the farmer's own Bt level, indicating that the results in the main regression are robust and that there is little to no peer learning.

Different behavioral responses. — Descriptive data can be useful in exploring other potential behavioral responses of farmers to learning (besides switching) where there is insufficient data to test the hypothesis empirically. 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.[42] The surveys ask the farmers in 2013 and 2014 to list the type of supplier they used and the time it took to reach the supplier using normal transport. 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.[43] Based on this, I estimate that two-thirds of the farmers did not change their supplier. **Figure A2a** in the appendix shows no correlation between the change in supplier and Bt content. It appears unlikely that learning occurred and drove supplier switching.

Another possibility, albeit remote, is that the farmers uncovered Bt content and reacted by exiting cotton production altogether. There is insufficient information to control for factors that can influence the decision to exit, but, qualitatively, farmers who exited cite predominantly environmental reasons in the surveys as shown in **Figure A2b**. Additionally, **Figure A2c** relates the 543 farmers who cultivated cotton in 2013 and for whom we have information about whether they cultivated again in 2014, to Bt content in 2013. There is no difference in the mean Bt gene expression between the group that exited and the one that remained.

Measurement error. — It is possible that the behavioral models and empirical specifications are accurate but that insignificance is due to attrition bias from

[42]The others did not answer the questions necessary for me to construct this best estimate guess.
[43]I assume significant change if the commuting time more than doubled or less than halved.

measurement error in the explanatory variable of interest. 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, and values may differ between plants as well as between leaf and boll. Whereas leaf values seem to be significantly correlated between the two plants for each farmer (with a correlation of about 0.7), the boll values seem to be much less correlated. Therefore, it is possible that that sample does not accurately represent the "true" Bt Level for the variety grown by that farmer (only obtainable with certainty if we measured the toxicity of all the plants per plot, i.e. destroy the farmer's crop).

In **Table A2** I explore this limitation by redefining the Bt content variable to reduce possible measurement error and rerunning the main regression in Column 3 of Table 4. In Column 1, I run the regression on the sample using an average of the leaf values only instead of leaf and boll, since leaf values are more strongly correlated with each other. In Column 2, I still use the leaf values but with one value as an instrument for the other.[44] In Column 3 I create a small subsample (72 observations) where the leaf values from both random plants were very similar to each other (ratio of values was close to one) and estimate the model for that subsample.

As shown in the table, the results do not change in any significant way. Therefore, while the size of the biophysical sample per farmer is very small and measurement error may certainly exist and contribute to inflating the standard errors, it does not appear that Bt content, even measured in more restrictive ways, impacts seed choice as we would expect if the farmers were uncovering this biological characteristic from cultivation experience. 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.

[44]The idea is that this will eliminate correlated noise or measurement error; a similar approach is used in Ashenfelter and Krueger (1994) to estimate returns from schooling using data on twins.

Additional robustness checks. — To further check the robustness of the main regression (Table 4, effect of Bt on variety change), I focus on Column 3 and introduce in **Table A3** (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.[45] I also (iv) reestimate 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. The results show that the conclusions are robust.

To check the effects of clustering the dependent variable in Table 6 (effect of perceptions on variety change), I estimate Column 3 as an ordered logit, with perceptions taking all three values (poor, moderate, and very good) and ordered as such. The results in **Table A4** demonstrate the robustness of the findings to the specification of the dependent variable.[46]

The above checks introduce a few alternative specifications. To test robustness much more widely, I utilize specification curve analysis, where all the plausible and valid controls are combined and/or omitted in hundreds or thousands of different ways, and the coefficient of interest estimated and plotted across all specifications (Simohnson et al, 2015; Rohrer, 2018; Orben et al, 2019).

Figure A3 demonstrates the robustness of the results on the effect of Bt on variety change. The controls for Equation (12) - farmer characteristics, planting history, prices, input use, and district fixed effects - are combined in different ways, leading to around 8,000 different specifications. The figure plots the effect of Bt content on variety change in each of these specifications; insignificant coefficients are black while significant ones are red. The majority of estimates are very close to zero, between -1% and +1%, and in *not one* of these specifications is the effect

[45]I omit these dummies from the presentation.

[46]The interpretation of the coefficients is no longer straightforward but the significance or lack thereof of coefficients is the same. Sowing time and district name are accounted for but omitted from the table.

of Bt on variety change significant. Therefore, this result is extremely robust.

Similarly, **Figure A4** plots the estimated effect of having very good perceptions (relative to poor) on variety change across the thousands of possible specifications. It shows that most estimates are negative and large, from -10% to -20%, confirming the result that farmers are less likely to switch varieties when their perceptions are more positive.[47] **Figure A5** demonstrates the robustness of the result, across all plausible specifications, that Bt content does not drive perception formation,[48] while **Figure A6** demonstrates the robustness of the result that Bt content does not inform pesticide use.

VIII. 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 elicit information, to assess which technologies maximize production efficiency, to know how to best adapt new technical and organizational skills, and to evaluate 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

[47] Only specifications where district controls are omitted are insignificant; this is a poor emission as district specific cultivation attitudes likely affect both perception formation and seed selection habits.

[48] Only specifications where district controls are omitted are negative. This is a poor emission as district effects partly control for exogenous pest intensity, in addition to informing cultivation attitudes.

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 an econometric specification to test for learning, since whether agents can learn and redress information problems is ultimately an empirical question. Using a rich and underutilized dataset, I apply the empirical exercise 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.

REFERENCES

- Ali A and Abdulai A. (2010) "The adoption of genetically modified cotton and poverty reduction in Pakistan." *Journal of Agricultural Economics*, Vol. 61 pp. 175-192.

- 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*, Vol. 42, pp. 1583-1594.

- 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*; Vol. 102, pp. 1632-1636.

- Ashenfelter O and Krueger A. (1994) "Estimates of the economic return to schooling from a new sample of twins." *American Economic Review*, Vol. 84 No. 5, pp. 1157-1173.

- Auriol E and Schilizzi S. (2015) "Quality signaling through certification in developing countries" *Journal of Development Economics*, Vol. 116 pp. 105-121.

- Bardhan P and Udry C. (1999) *Development Microeconomics.* Oxford: Oxford University Press.

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

- Braverman A (1985). "Rural credit markets and institutions in developing countries." *World Development.* Vol 14.

- Besley T (1994). "How do market failures justify interventions in rural credit markets?" *World Bank Research Observer.* Vol 9 pp. 27-47.

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

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

- Darby M and Karni E. (1973) "Free competition and the optimal amount of fraud." *The Journal of Law and Economics*, Vol. 16 pp. 67-88.

- Dean M. (2019) "Rational inattention lecture." Department of Ecnomics, Columbia University, personal website.

- Dosi, G. ed. (1988). *Technical Change and Economic Theory.* Laboratory of Economics and Management (LEM) Sant'Anna School of Advanced Studies, Pisa.

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

- 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*, Vol. 101 pp. 526-55.

- Emerick E, Janvry A, Sadoulet A, and Dar MH. (2016). "Enhancing the diffusion of information about agricultural technology." Working paper.

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

- Foster AD and Rosenzweig M. (2010) "Microeconomics of technology adoption." *Annual Review of Economics*, Vol. 2 pp. 395-424.

- 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.

- Girard T and Dion P. (2010) "Validating the search, experience, and credence product classification framework." *Journal of Business Research*, Vol. 63 pp. 1079-1087.

- 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.*

-ISAAA (2017). “Global status of commercialized Biotech/GM crops in 2017.” *ISAAA Brief No. 53*.

- Khan, M. (2010) “Learning, technology acquisition, and governance challenges in developing countries.” Unpublished manuscript, SOAS Economics Dpt.

- Khan, M. (2012) “Beyond good governance: An agenda for developmental governance.” Unpublished manuscript, SOAS Economics Department.

- Khan, M. (2013) “Technology policies and learning with imperfect governance.” In Stiglitz and Lin (eds) *The industrial policy revolution I: The role of government beyond ideology*, London: Palgrave, pp. 79-115.

- Kouser S and Qaim M. (2014) “Bt cotton, damage control and optimal levels of pesticide use in Pakistan.” *Environment and Development Economics*, Vol. 19 pp. 704-723.

- 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*, Vol. 68, pp. 366-385, Vol 73 pp. 185-213.

- Ma X and Shi G. (2015) “A dynamic adoption model with Bayesian learning: an application to US soybean farmers.” *Agricultural Economics* Vol 46 pp. 25-38.

- Maertens A, Michelson H, and Nourani V. (2018) “How do farmers learn from extension services? Evidence from Malawi”. *SSRN Electronic Journal*.

- Munshi K. (2004) “Social learning in a heterogenous population: technology diffusion in the Indian Green Revolution.” *Journal of Development Economics*.

- Murphy D. (2017) “Underground knowledge: soil testing, farmer learning, and input demand in Kenya.” Unpublished.

- Nelson P. (1970) “Information and consumer behavior.” *Journal of Political Economy* Vol. 78 pp. 311-29.

- Orben A and Pryzbylski A. (2019) “The association between adolescent well-being and digital technology use.” *Nature Human Behavior*. Vol 3 pp. 173-182.

- 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*, Vol. 3 pp. 44-56.

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

- Rohrer JM. (2018) "Run all the models! Dealing with data analytic flexibility." *Association for Psychological Science Observer*: March 2018 ed.

- Simohnson U, Simmons J, and Nelson L. (2015) "Specification curve: descriptive and inferential statistics on all reasonable specifications." *LSE Working Paper, STICERD.*

- Sims C. (2003) "Implications of rational inattention." *Journal of Monetary Economics*, Vol. 50 pp. 665-690.

- 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*, Vol. 18 pp. 98-112.

- 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.

- Stiglitz, J. (1989). "Markets, market failures, and development." *American Economic Review* Vol. 79 pp.197-203.

- 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.*

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

- USDA Foreign Agricultural Services (2019). "Pakistan: Cotton and Products Annual."

- 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* Vol. 46 pp. 183-192.

Appendix

A1. Deriving Equation (9)

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

A2. Translating Bt improvement into monetary gains

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*, like above, 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.[49]

1) Calculating the size of the industry is straightforward. According to the Pakistani government, cotton production in Pakistan in 2014-2015 totalled 13,960,000 bales,[50] equivalent to about 2.37 billion kg.[51] 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. [52]

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

This can be used to extrapolate differences in lethality based on Bt content. For example, an improvement in mean Bt content from 0.88 $\mu g/g$ to 0.97$\mu g/g$ would raise killing effectiveness from 70% to 75%. The question

[49] I assume that p can vary $\in (0, 1)$, and that σ can also vary in $(0, 1)$. Though (unlike p) σ does not need to be bounded at 1, it is unlikely that it exceeds this amount; Bt content usually varies between 0 and 3 $\mu g/g$ and even in a small sample like mine the standard deviation is ≈ 0.57.

[50] `http://www.finance.gov.pk/survey/chapters_18/02-Agriculture.pdf`

[51] In Pakistan, one bale of cotton is equivalent to 170 kg.

[52] I cannot make assertions about the total size of the industry because I do not have data about the distribution of end buyers and the prices that these different buyers pay.

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.[53][54]

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. The figure below shows monetary gains, on the vertical axis, in millions of USD for $\alpha \in [0, 0.2], p \in [0, 1], \sigma = 0.57$. For example, at $\alpha = 0.1$ 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 $0.926\mu g/g$. 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 graph. [55]

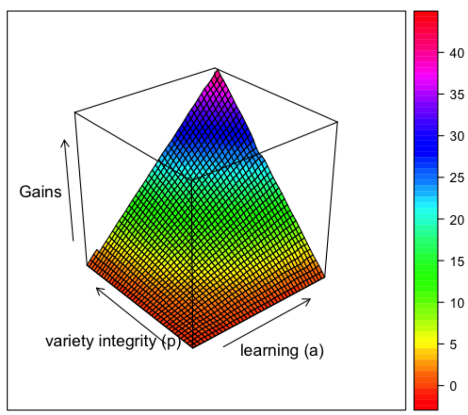


Figure A1. Monetary gains in the market

Note: Figure A1 plots market-wide monetary gains in millions of USD as a function of both the learning parameter a (axis from 0-0.20) and variety integrity (axis from 0-1).

[53] This is on the assumption that increased output is absorbed without reducing prices.

[54] This is a lower bound estimate. The data, from Spielman et al (2015), shows that moving from under 0.88 to over that threshold improves yield by 50%. Calculating gains this way - by examining the share of farmers that move below the threshold to above it - for the example above (Bt from 0.88 to 0.97) results in similar but somewhat higher estimated gains of 3.2%, or 44 billion USD.

[55] With higher σ the graph would tilt further up, generating more gains for any combination of learning and variety integrity.

A3. Controls

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 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 cotton season.

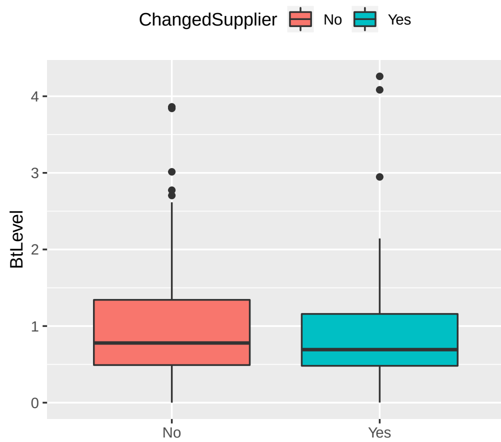
A4. Alternative explanations

Table A1—Including a measure of peer learning

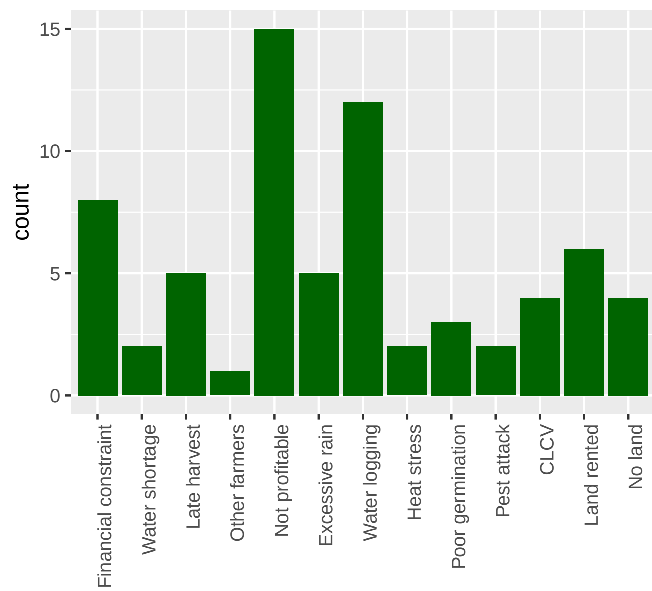
	Dependent variable:			
	CHANGED			
	(1)	(2)	(3)	(4)
Bt level	0.001	−0.011	−0.010	−0.010
	(−0.101, 0.103)	(−0.113, 0.090)	(−0.108, 0.088)	(−0.104, 0.083)
Other-variety neighbors	0.133	0.187	0.173	0.302
	(−0.268, 0.534)	(−0.227, 0.600)	(−0.206, 0.553)	(−0.079, 0.683)
Education		0.008	0.017***	0.016**
		(−0.005, 0.021)	(0.004, 0.031)	(0.003, 0.030)
Farming experience			0.009***	0.008**
			(0.002, 0.016)	(0.001, 0.015)
Years variety grown			−0.092***	−0.101***
			(−0.157, −0.028)	(−0.163, −0.038)
Years Bt grown			−0.062***	−0.059***
			(−0.103, −0.022)	(−0.101, −0.018)
Land owned			−0.010**	−0.011***
			(−0.018, −0.001)	(−0.018, −0.003)
Purchase price (seed)		−0.00001	−0.0001	−0.00002
		(−0.001, 0.001)	(−0.001, 0.0004)	(−0.001, 0.001)
Selling price (cotton)		−0.029*	−0.027	−0.027
		(−0.064, 0.005)	(−0.062, 0.009)	(−0.064, 0.010)
Irrigation				−0.0001
				(−0.0002, 0.00004)
Fertilizer				−0.0005
				(−0.002, 0.001)
Seed amount				0.039***
				(0.013, 0.065)
Labor				0.001*
				(−0.0001, 0.001)
Pesticide				−0.00000
				(−0.0001, 0.00004)
District FE	Yes	Yes	Yes	Yes
Observations	233	233	233	233
R^2	0.126	0.139	0.233	0.275
Adjusted R^2	0.066	0.067	0.153	0.180
Residual Std. Error	0.474 (df = 217)	0.474 (df = 214)	0.452 (df = 210)	0.444 (df = 205)

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

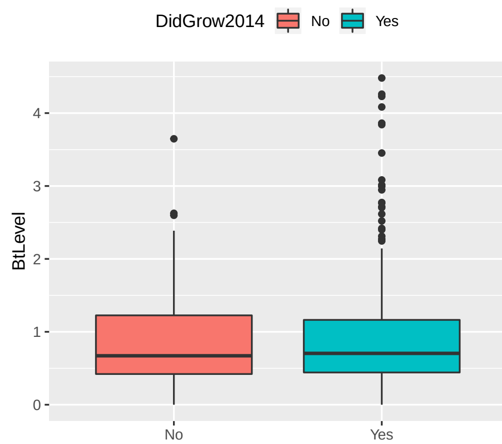
Note: Table A1 includes a rough measure of peer effects: the average Bt of farmers who cultivated a different variety in 2013. If the variable captures real peers, then social learning implies the coefficient on this variable would be positive, with farmers more likely to switch varieties when they learn that the Bt of farmers who purchased a different variety is high. Across specifications, the coefficient is close to zero and we cannot reject the null, suggesting there is no learning from peers. The coefficient on learning from own Bt is also robustly null.



(a) Supplier change



(b) Reasons they exited



(c) Correlation between Bt level and exit

FIGURE A2. EXCLUDING ALTERNATIVE BEHAVIORAL RESPONSES

Note: Figure A2 considers behavioral responses to learning besides variety switching. Panel a shows it is unlikely that farmers reacted to low Bt level by switching suppliers in the next year. Panel b 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; Panel c complements this by showing no correlation between the farmer's Bt level and the decision to exit production.

Table A2—Accounting for measurement error

	Dependent variable:		
	Changed variety		
	(1)	(2)	(3)
BtLevel - leaf average	−0.019		
	(−0.082, 0.045)		
BtLevel - instrumented leaf		−0.024	
		(−0.163, 0.114)	
BtLevel - correlated leaf only			−0.020
			(−0.120, 0.080)
Education	0.015**	0.015**	0.024**
	(0.003, 0.026)	(0.002, 0.028)	(0.002, 0.047)
Farming experience	0.008***	0.009***	0.019***
	(0.002, 0.014)	(0.003, 0.015)	(0.008, 0.031)
Years variety grown	−0.073***	−0.076***	−0.101*
	(−0.127, −0.019)	(−0.133, −0.019)	(−0.202, 0.0003)
Years Bt grown	−0.040**	−0.043**	−0.130***
	(−0.075, −0.006)	(−0.078, −0.007)	(−0.181, −0.078)
Land owned	−0.007*	−0.007*	−0.015*
	(−0.014, 0.001)	(−0.015, 0.001)	(−0.031, 0.002)
Seed purchase price	−0.0003	−0.0004	−0.001*
	(−0.001, 0.0002)	(−0.001, 0.0002)	(−0.002, 0.0002)
Selling price	−0.024	−0.025	−0.030
	(−0.054, 0.007)	(−0.057, 0.007)	(−0.086, 0.026)
District FE	Yes	Yes	Yes
Observations	301	290	72
R^2	0.284	0.269	0.607
Adjusted R^2	0.202	0.181	0.379
Residual Std. Error	0.444 (df = 269)	0.450 (df=258)	0.384 (df = 45)

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

Note: Table A2 demonstrates the results from reconstructing the Bt variable to reduce measurement error and reestimating the effect of Bt content on variety change. In Column 1 I reconstruct 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. In Column 2 I use one leaf value as an instrument for the other to eliminate (the correlated) measurement error. In Column 3 I keep Bt content as the average of the leaf and boll values but apply it only to a limited set of observations (72) where the two leaf values are almost identical. In all, Bt still has an insignificant effect on variety change.

A5. *Further robustness checks*

TABLE A3—ALTERNATIVE SPECIFICATIONS TO CHECK EFFECT OF BT LEVELS ON VARIETY CHOICE

	Dependent variable:			
	Changed variety			
	(1: LPM)	(2: LPM)	(3: LPM)	(4: Logit)
Bt level (standardized)	0.018	−0.007	0.001	−0.023
	(−0.077, 0.114)	(−0.071, 0.057)	(−0.061, 0.063)	(−0.364, 0.310)
Bt level squared	−0.008			
	(−0.038, 0.022)			
Education	0.014**	0.014**	0.015**	0.072**
	(0.001, 0.026)	(0.002, 0.026)	(0.004, 0.027)	(0.011, 0.137)
Bt:Education		0.003		
		(−0.008, 0.014)		
Farming experience	0.008***	0.008***	0.007**	0.041***
	(0.002, 0.013)	(0.002, 0.014)	(0.002, 0.013)	(0.013, 0.071)
Years variety grown	−0.075***	−0.074***		−0.366***
	(−0.130, −0.020)	(−0.129, −0.020)		(−0.635, −0.114)
Years Bt grown	−0.040**	−0.040**	−0.040**	−0.201**
	(−0.074, −0.005)	(−0.075, −0.006)	(−0.075, −0.006)	(−0.388, −0.024)
Land owned	−0.007*	−0.006*	−0.007*	−0.032**
	(−0.014, 0.001)	(−0.014, 0.001)	(−0.014, 0.001)	(−0.064, −0.0001)
Purchase price (seed)	−0.0003	−0.0004	−0.0004	−0.002
	(−0.001, 0.0002)	(−0.001, 0.0002)	(−0.001, 0.0001)	(−0.004, 0.0007)
Selling price (cotton)	−0.023	−0.023	−0.016	−0.146*
	(−0.053, 0.008)	(−0.053, 0.008)	(−0.046, 0.014)	(−0.319, 0.021)
District FE	Yes	Yes	Yes	Yes
Observations	301	301	301	301
R^2	0.284	0.284	0.310	
Adjusted R^2	0.199	0.199	0.216	
Residual Std. Error	0.445 (df = 268)	0.445 (df = 268)	0.440 (df = 264)	

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

Note: Table A3 introduces different specifications to Column 3 in Table 4 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. In all specifications, we still cannot reject a null of no effect of Bt on variety choice.

Table A4—Ordered logit to check effect of Bt on perception formation

	Dependent variable:
	Perception (Ordered)
	Logit
Bt level (standardized)	0.037
	(−0.246, 0.321)
Education	0.036
	(−0.020, 0.093)
Farming experience	0.003
	(−0.023, 0.030)
Years variety grown	0.009
	(−0.227, 0.245)
Years Bt grown	−0.034
	(−0.203, 0.134)
Seed price	−0.001
	(−0.003, 0.002)
District and sowing-time FE	Yes
Observations	301

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

Note: Table A4 reestimates the effect of Bt on farmer perceptions by including all three levels of farmer perceptions in the dependent variable, with an ordered logit. This is in contrast with the results in Table 6, which use a linear probability model and cluster perceptions into a binary 'poor/moderate' versus 'very good' variable. As shown, the basic results are the same: Bt has no discernable effect on farmer perceptions of bollworm resistance.

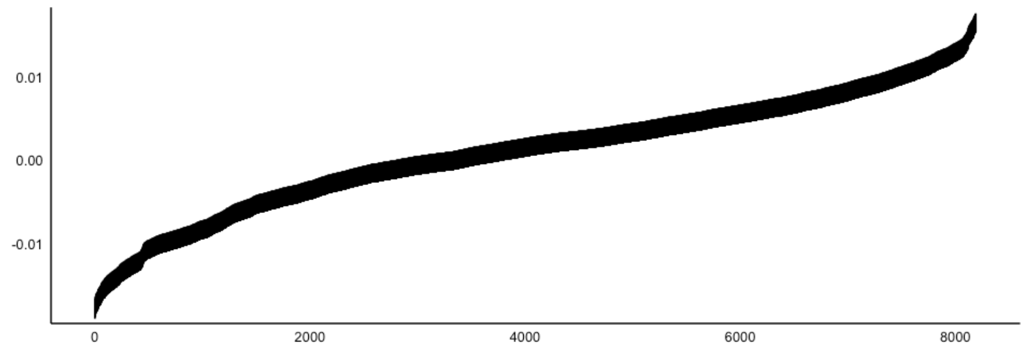


Figure A3. Specification curve analysis: Effect of Bt on variety change

Note: The x-axis is the number of specifications, and the y-axis is estimate of the effect of a one standard deviation increase in Bt on the probability of variety change in each specification. *Black* denotes an insignificant effect while *red* denotes a significant effect, at the 5% significance level. The specifications combine all the possible controls (farmer characteristics, planting history, prices, input use, and district controls) in thousands of different ways. In all of these specifications the effect of Bt on the probability of variety change is close to zero and insignificant.

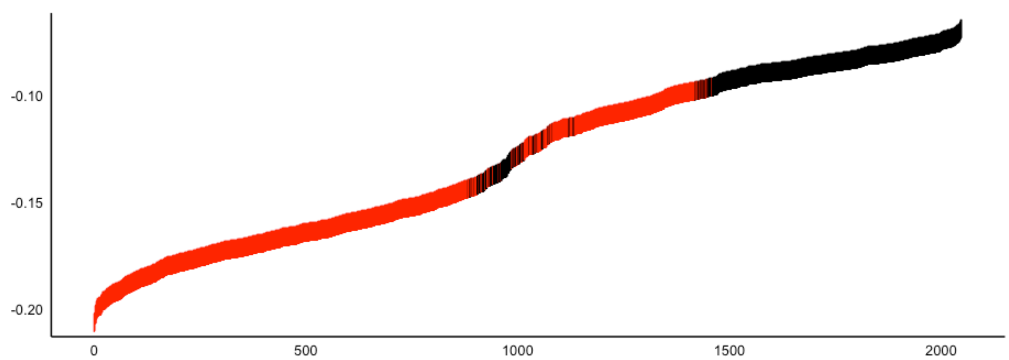


Figure A4. Specification curve analysis: Effect of perception on variety change

Note: The x-axis is the number of specifications, and the y-axis is estimate of the effect of viewing perception resistance as very good, relative to poor, on the probability of variety change. *Black* denotes an insignificant effect while *red* denotes a significant effect, at the 10% significance level. The specifications combine the possible controls (farmer characteristics, planting history, prices, pesticide use, yield, and district controls) in thousands of different ways. In the majority, the effect of very good perceptions on the probability of variety change is negative and significant. The exceptions in the upper right corner are the specifications that omit district controls, which is not a valid omission since district-specific cultivation attitudes are likely correlated with both perceptions and seed selection behaviors.

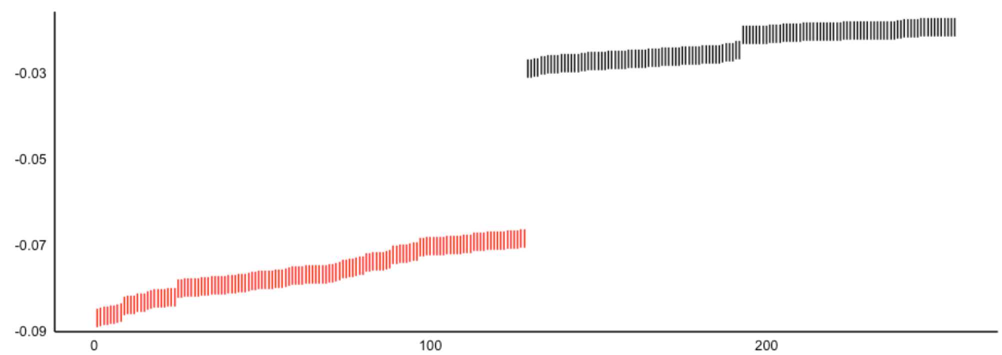


Figure A5. Specification curve analysis: Effect of Bt on perceptions

Note: The x-axis is the number of specifications, and the y-axis is estimate of the effect of a one standard deviation increase in Bt on the probability of viewing resistance as very good instead of poor/moderate. *Black* denotes an insignificant effect while *red* denotes a significant effect, at the 5% significance level. The specifications combine the controls (farmer characteristics, planting history, seed price, pesticide use, and sowing time and district controls) in hundreds of different ways. In the specifications that exclude district controls (the lower left corner), the estimated coefficient is negative and significant. However, in the more plausible specifications that include district controls (upper left), the estimated effects of Bt on resistance perceptions are closer to zero and insignificant, consistent with the regressions.

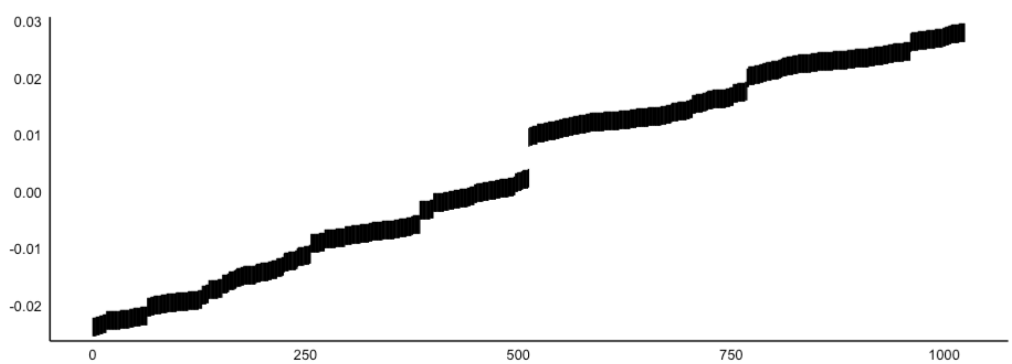


Figure A6. Specification curve analysis: Effect of Bt on pesticide use

Note: The x-axis is the number of specifications, and the y-axis is estimate of the effect of a one standard deviation increase in Bt on the use of pesticide per acre cultivated. *Black* denotes an insignificant effect while *red* denotes a significant effect, at the 5% significance level. The specifications combine the possible controls (farmer characteristics, planting history, area cultivated, seeds sowed per acre, and controls for soil type, sowing time, and district) in thousands of different ways. In all specifications the estimated effect of Bt on pesticide use is close to zero and insignificant.