Received: 12 December 2020 | Accepted: 11 April 2021

DOI: 10.1111/ajae.12276

ARTICLE

Uncertainty and learning in a technologically dynamic industry: Seed density in U.S. maize

Edward D. Perry[1] | **David A. Hennessy**[2] | **GianCarlo Moschini**[2]

[1]Department of Agricultural Economics, Kansas State University, Manhattan, Kansas, USA

[2]Department of Economics and the Center for Agricultural and Rural Development, Iowa State University, Ames, Iowa, USA

Correspondence
Edward D. Perry, Department of Agricultural Economics, Kansas State University, Manhattan, KS, USA.
Email: edperry@ksu.edu

Abstract

The large and sustained yield gains achieved since the introduction of maize hybrids in the 1930s (about 1.8 bushels per acre per year) have been accompanied by a remarkably parallel and steady increase in seeding density. This increase occurred in an environment characterized by rapid technological innovation, including genetic engineering, and commercial hybrid varieties with short life cycles. An important question, then, is whether and how breeders and farmers have learned about the optimal planting density. In this paper, we use unique and detailed U.S. farm-level data, consisting of more than 400,000 planting choices from 1995–2016, to assess the nature of learning about seeding density. Importantly, we control for unobserved confounders through both hybrid and farm-level fixed effects. We find that the variance in planting rates for a given hybrid has decreased over time, and that farmers tend to plant a given variety at higher rates over time. This is consistent with Bayesian learning in which risk-neutral farmers possess priors consistently below the true optimal rate. We cast doubt on risk aversion as a credible explanation for this finding by analyzing the contrasting evolution of soybean planting rates (a crop with exogenously different agronomic determinants of seed density). We interpret our results as evidence of *inertia*: the initial bias in maize farmers' priors is tilted towards the optimal planting rates of varieties planted in the past. One implication of the finding that farmers historically underinvested in seeding rates is that eliminating this tendency could result in productivity gains.

KEYWORDS

Bayesian learning, inertia, maize hybrids, productivity, seeding density, uncertainty

JEL CLASSIFICATION
D22, D80, Q10, Q12

1 | INTRODUCTION

One of the most significant agricultural advances of the 20th century was the discovery of hybrid vigor in maize. The subsequent commercialization of maize hybrid varieties in the 1930s was met with widespread adoption by U.S. farmers (Griliches, 1957), and over the ensuing 80 years U.S. average maize yields increased at the impressive rate of about 1.8 bu/acre/year, from roughly 20 bu/acre in the mid-1930s to more than 170 bu/acre in 2016, a nearly eight-fold increase. A large body of research has shown that these yield gains were the result of complex interactions between genomic advances due to plant breeding and improved agronomic practices, including enhanced nitrogen fertilizer and pesticide use (Duvick, 2005). Whereas the exact contribution of each factor remains a matter of debate, one unambiguous statistical fact stands out: Virtually all observed yield increases have been accompanied by increasing planting density (i.e., plants per acre). For example, from the 1960s to the present, mean plant density more than doubled, from 14,000 to 30,000 plants per acre. There is a strong agronomic basis for this relationship, discussed further below. In any given year, however, the actual optimal seed density for any given variety is, even having controlled for the specific soil and climatic conditions faced by heterogeneous farmers, highly uncertain. Farmers have a large set of seed varieties to choose from and, due to the continual introduction of new varieties, there is also a rapid turnover in this set. The dynamic and complex nature of the industry means that farmers' seed density choice is a difficult problem, one that likely entails errors as well as learning. Yet, this decision context has remained largely unexplored to date.

Learning is critical in environments where new, complex technologies are frequently introduced (Foster & Rosenzweig, 2010). The presence of learning has been documented in many contexts, but its importance has been particularly striking in agriculture. Unlike many other industries, technological progress is mostly exogenous to the agricultural production sector, being rooted in research and development investments undertaken by public laboratories or input-supplying firms (Clancy & Moschini, 2017). As a result, farmers continuously face adoption choices vis-à-vis a flow of technological innovations they do not control. A growing literature has sought to assess whether and how farmers learn about new technologies (Bold et al., 2017; Conley & Udry, 2010; Duflo et al., 2011; Emerick & Dar, 2021; Foster & Rosenzweig, 1995; Hanna et al., 2014). This work has dealt primarily with developing countries, where adoption of new technologies can be extremely slow, and insofar as it has considered optimal input decisions it has typically focused on fertilizer use (Suri, 2011) and variety adoption (Matuschke & Qaim, 2009). A common thread in these studies is that farmers not only learn but that this learning can be characterized by initial *underinvestment*: Producers appear to use too little, rather than too much, of an input (Foster & Rosenzweig, 2010). For example, Duflo et al. (2008, 2011) find that Kenyan farmers underinvest in profitable fertilizer investments.

In this paper, we study the importance of uncertainty and learning for U.S. farmers' decisions about the optimal planting density (seeds/acre) for maize hybrids. What makes this context unique is the rapid pace of both product turnover and technological change. As we document below, the commercial life cycle of a maize hybrid variety is typically less than seven years, and in any given year there are hundreds of horizontally differentiated varieties to choose from. Thus, U.S. farmers have had to continually re-evaluate the optimal planting rate in a highly complex and changing environment. To conduct our analysis, we draw on a large, unique U.S. farm-level dataset of more than 400,000 maize planting choices. The data spans the period 1995–2016 and contains information on the specific maize hybrid planted, seed price, and farmer-chosen seeding density. These data permit us to empirically assess whether farmers' planting rate choices were consistent with testable

predictions that emerge from a model in which learning is present. In addition, we assess whether this learning was characterized by initial over or under investment in seed density.

To guide the empirical analysis, we first develop a Bayesian learning model in which risk-neutral farmers choose the seeding density that maximizes expected profit. This model is in the tradition of the target-input models employed in the prior literature (Foster & Rosenzweig, 2010). Consistent with agronomic evidence (e.g., Assefa et al., 2016), we specify maize yields as quadratic in seeding rates. Uncertainty arises because farmers have imperfect knowledge about this function and, in turn, the optimal planting rate. Over time, as farmers (individually and collectively) gain experience with a particular variety, they update their expectations of the optimal rate. We show that this learning process is characterized by two important, empirically relevant features. First, the presence of learning implies that the variance in farmers' beliefs about the optimal rate decreases over time, which in turn implies that the (conditional) variance in chosen planting rates decreases over time. Second, farmers' prior beliefs about the optimal rate can be below or above the true optimal rate. Given a population of risk-neutral farmers who are unbiased in their priors, we should not observe positive (or negative) within-variety planting rate trends. Evidence to the contrary suggests that learning with a biased prior has occurred.

To assess these two implications, we estimate a series of linear models in which farmer-specific planting rates are regressed on the hybrid's commercial age, the ratio of the seed price to the expected maize price, and also variety and farmer fixed effects. Because variety fixed effects are included, the commercial age coefficient captures whether, on average, farmers' priors were below or above the truth. To assess whether there was a decrease in farmers' perception errors about planting rates over time, we estimate a second-stage regression in which the squared predicted residuals from the first stage are regressed on the commercial age variable (as well as variety and farmer fixed effects). Importantly, the inclusion of farmer fixed effects rules out time-invariant unobserved heterogeneity as a driver of the estimated impacts.

The empirical analysis yields two primary findings. First, there is a statistically significant reduction in the estimated variance of planting rates in the years after a variety enters the market. Each additional year of commercial availability is associated with a decrease in variance of approximately 4%. Thus, there is clear evidence of learning in the industry. Second, variety-specific mean planting rates increase significantly with the industry's experience with the variety. On average, one additional year of commercial availability is associated with an increase in planting rates of about 222 kernels per acre, equivalent to about 0.75% of the national mean planting rate. This suggests that farmers' initial priors are systematically below the true optimal rate; that is, they exhibit chronic underinvestment in seeding rates. We cast doubt on risk aversion as a plausible explanation for this finding by showing that planting rates for soybeans, over the same period and for essentially the same set of farmers, have moved in the opposite direction (consistent with agronomic advice for this crop). Thus, we interpret our findings as evidence of *inertia*: The initial bias in farmers' priors is tilted toward the optimal planting rates of varieties planted in the past.

To illustrate the importance of our estimated results, we investigate two implications of learning for the industry. We first simulate planting rate counterfactuals in which the amount of learning is adjusted. We show that planting rates would have still increased at a similar rate over time, but the path would have shifted up by about 5.7% per year had farmers been initially endowed with more information. Using the simulated planting rate counterfactuals, we then consider how yield trends would have been affected by different amounts of learning. This exercise predicts that, in the states of Iowa and Illinois, maize yields would have been lower by about 1 bu/acre per year had learning not occurred. Assuming a maize price of $4/bu, this amounts to a $100 million gross revenue impact per year in these two states.

This paper contributes to the literature on learning in an agricultural context in several ways. As noted, no existing work addresses whether and how farmers learn about the optimal seed density. Most previous work that has assessed agricultural learning, primarily in developed countries, has typically done so with experimental data. Although there are advantages to using experimental data,

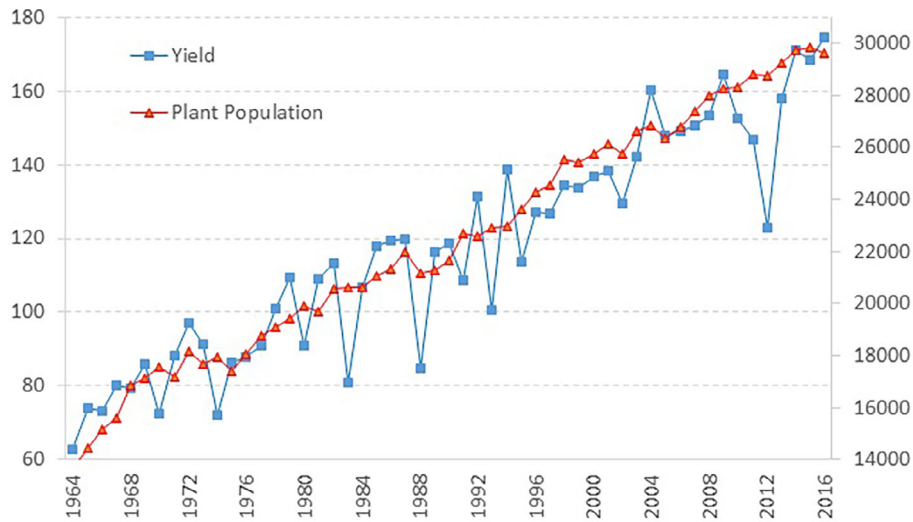

FIGURE 1 U.S. maize yield (left, bu/acre) and plant population (right, plants/acre): 1964–2016. Note: Plant population data is for the top 8 maize producing states. Source: USDA

well-known limitations include lack of external validity. Here we leverage a unique and large dataset of actual farm-level choices across a relatively long-time frame, permitting us to assess whether learning has occurred "in the field." Our results also demonstrate an important link between learning and food production. Maize is the most important field crop in the United States and is one of the world's key food commodities. Our findings highlight a distinctive feature of productivity growth in maize. Because the life cycle of new maize hybrids is relatively short, farmers are constantly "catching up" to the correct, typically higher, planting rate. This suggests there may be significant unrealized returns from additional investments into promoting learning on and reducing uncertainty about optimal planting rates.

2 | BACKGROUND

The link between observed maize yield increases and plant density is illustrated in Figure 1.[1] From the mid-1960s to the present, yields more than doubled (from about 75 bu/acre to 175 bu/acre), while maize plant populations also roughly doubled (from about 14,000 plants/acre to nearly 30,000 plants/acre). Thus, remarkably, yield *per plant* is only slightly greater now than it was over 50 years ago.

The fundamental agronomic reason for the observed positive relationship between plant populations and yields was demonstrated in a series of experiments in which new and old maize hybrids were planted at different densities (Duvick, 2005; Russell, 1991). These experiments revealed that at low densities—where water stress and overcrowding were non-limiting factors—the yield advantage of new hybrids was relatively small. At high densities, however, the yield gap was substantial. Thus, an essential attribute of newer hybrids is that they possess higher tolerance to crowding, thereby permitting higher planting densities and a higher number of harvested maize ears per acre (Tokatlidis & Koutroubas, 2004). The relationship between density and yields may be even stronger for genetically engineered (GE) varieties, introduced and widely adopted since the mid-1990s (Chavas et al., 2014).

[1]It is important to distinguish between two closely related but distinct concepts, "plant population" and "planting rate" (also referred to as "seeding rate"). Plant population is the number of plants per acre that are standing at the end of the growing season, whereas the planting rate is the number of planted seeds per acre. Our data and analysis pertain to seeding density (planting rate), whereas the USDA data used for Figure 1 relate to plant population. Planting rate trends for the state of Iowa can be found in Figure 2 below.

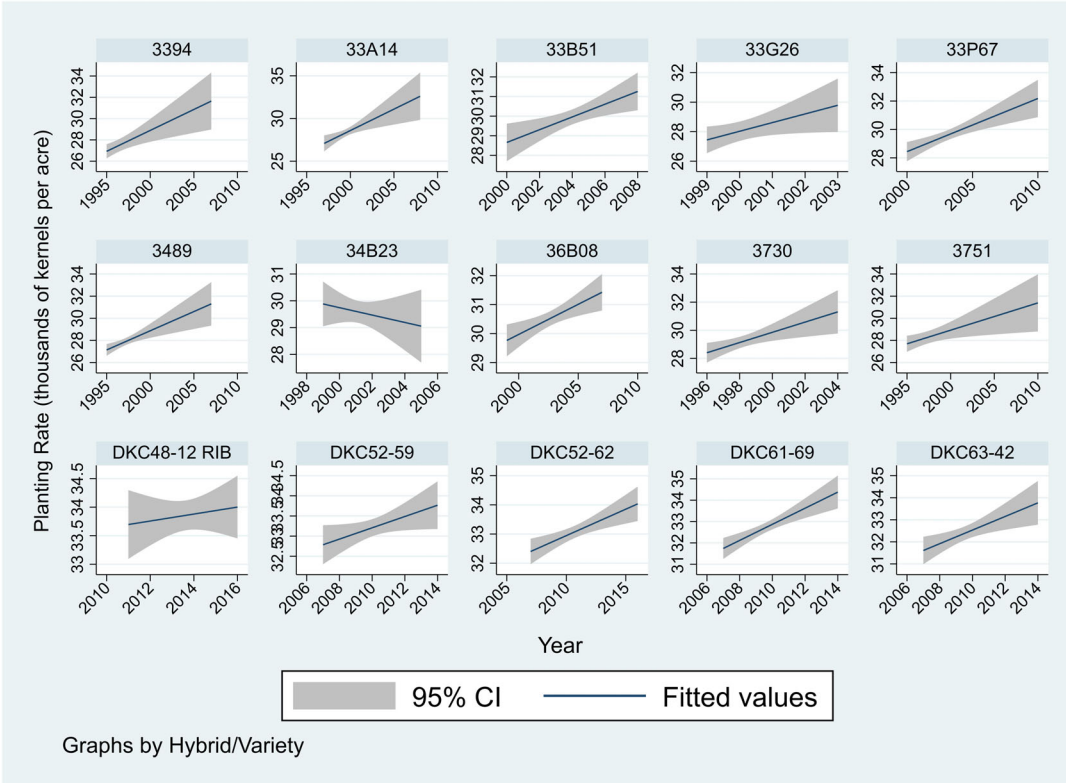

FIGURE 2 Planting rate trends for selected most popular hybrids in Iowa, 1995–2016

Plant density is particularly important for maize. Most other cereal crops can respond to environmental constraints (e.g., water and nutrient availability) by adjusting the number of productive tillers (the reproductive shoots that grow from the plant), whereas modern maize varieties cannot effectively do so (Assefa et al., 2016). Furthermore, the optimal plant density is not simply a genetic attribute of a hybrid variety; it is also heavily influenced by local environmental conditions and their interactions with genotype. All of this raises an important question: To what degree and how rapidly have farmers been able to discover the optimal rates of newer hybrids for their own location? In other words, to what extent has learning played a role in the observed increase in planting rates and yield growth over time?

Previous studies have often employed a simple test for the presence of learning: When efficiency and/or productivity increases for a fixed technology over time, then learning is asserted to have occurred (e.g., Jovanovic & Nyarko, 1995). In our context, the natural analog would be to consider whether yields or profits have increased for a given hybrid over time. Such an indirect route is not feasible for us, however, because we do not observe yield or profits at the farm level. But we can conduct a more direct analysis of the hypotheses of interest. That is, similar to Conley and Udry (2010), we look directly at whether the input rate associated with a fixed technology changes over time.

Figure 2 provides some further motivating stylized facts. It illustrates *observed* planting rate trends for the 15 most widely planted hybrids in our sample. To avoid issues of heterogeneity across states, this figure uses data only for Iowa, the most important maize growing U.S. state. Out of the 15 hybrids, 14 had an increasing trend, and one hybrid had a statistically insignificant decreasing trend (hybrid 34B23)—all despite the apparent lack of other major changes that would have encouraged within-variety increasing rates. Indeed, in the empirical analysis below we show that these trends persist even

after a variety of controls have been put in place. In any event, farmers are clearly changing their input rates for a fixed technology over time, which suggests that learning may be occurring.

Our work is broadly related to the large literature on "learning by doing" in a production setting (Syverson, 2011). This concept, first articulated in Arrow's (1962) seminal paper, has been studied in contexts such as manufacturing (Argote & Epple, 1990), shipbuilding (Thompson, 2001), and automobile assembly (Levitt et al., 2013). One way to evince learning is to track productivity (efficiency) for a fixed technology over time. For example, Benkard (2000) found that the number of labor hours required to produce the Lockheed L-1011 TriStar had halved by the 30th plane and halved again by the 100th plane. A related method, used in Levitt et al. (2013), is to track a qualitative measure of learning (e.g., number of defects) over time. Alternatively, a more direct approach considers whether a chosen input level for a fixed technology changes as more experience is accumulated. For example, Conley and Udry (2010) investigate whether pineapple farmers adjusted their fertilizer use in response to both their own experience and that of their neighbors. Our approach follows this tack. By using a unique dataset of farmers' seed choices at the (very refined) variety level, and by including variety fixed effects in our model, our empirical results characterize whether and how farmers' use of a fixed technology evolves over time.

3 | MODEL

We assume that farmers choose the seeding rate that maximizes expected profit. The full model below will encompass the seed choices of many farmers, each choosing multiple varieties, and over several years. For clarity, though, consider first the problem faced by a single farmer for a given variety. We assume risk neutrality (the modeling framework and main results can be readily extended to the case of risk aversion, however) and presume that the farmer chooses the seeding rate to maximize expected profit. Experimental evidence from agronomic trials indicates that per-acre maize yields are approximately quadratic in seeding rates (Assefa et al., 2016). Without loss of generality, we can write a quadratic yield function as

$$y = b\left(\theta z - 0.5z^2\right) \quad , \quad \theta > 0,\ b > 0. \tag{1}$$

where z denotes seed density (amount of seed per acre). Note that, with this parameterization, yield is maximized at $z = \theta$. Hence, θ can be interpreted as what Assefa et al. (2018) term the "agronomic optimal" planting density (AOPD). The parameter θ is therefore, at least up to adjustments for input and output prices, a natural candidate for modeling how farmers learn about the optimal seed density.

From the farmer's perspective, the agronomic optimal density is a random variable, say $\widetilde{\theta}$. Let $\overline{\theta} \equiv E[\widetilde{\theta}]$ denote the expectation of the optimal density conditional on available information. Per-acre expected profit can then be expressed as

$$\pi = pb\left(\overline{\theta} z - 0.5z^2\right) - wz \tag{2}$$

where p denotes output price, and w denotes seed price. Maximizing the expected profit function in (2) yields the farmer's "economically optimal" seeding density:

$$z^* = \overline{\theta} - \frac{w}{pb} \tag{3}$$

This simple setup shows that the farmer's seed choice is a monotonic function of the expected agronomic optimal planting density. As farmers' information about the variety being planted improves, they will update their beliefs about $\overline{\theta}$ and adjust seed density, z^*, accordingly. The latter, of course, is an economic decision and must reflect variables relevant to the profit maximization problem other than the AOPD, such as the seed price w and the expected output price p. In any event, observations

about farmers' seed choices, z, provide a window into the nature of their learning process about θ. To gain more insights into such learning, we place more structure on the problem by developing a simple Bayesian learning model.

3.1 | Bayesian learning

To articulate a learning model suitable for the data at hand, we recognize that farms, denoted by the subscript i, are heterogeneous in land characteristics, and we consider several varieties, denoted by the subscript j. Learning takes place over time, each period tracked by the subscript t. When variety j first comes to market, at time $t = 0$, it is associated with the "ideal" but unknown planting density on farm i, denoted by θ_j^i. This ideal planting density can be decomposed into a variety-specific component θ_j, which is unknown and common across all farmers, and a farm-specific systematic component known to the farmer and denoted by α_i. One way to think of this framework is as a target input model along the lines of Foster and Rosenzweig (1995) and Jovanovic and Nyarko (1995). That is, we posit

$$\theta_j^i = \alpha_i + \theta_j \tag{4}$$

At periods $t = 1, 2, \ldots$, each farmer receives a signal s_{jt}^i, about the true planting density θ_j^i, where

$$s_{jt}^i = \theta_j^i + \varepsilon_{jt}^i \quad , \quad t = 1, 2, \ldots \tag{5}$$

The term ε_{jt}^i is an i.i.d. random variable, which captures the fact that the signal is noisy. Given this information, each farmer can update their beliefs, that is, form a posterior distribution for the random variable θ_j. Following previous literature (e.g., Erdem & Keane, 1996), we assume that both the prior and the noise have a normal distribution:

$$\theta_j^i \sim N\left(\overline{\theta}_{j0}^i, \sigma_{j0}^2\right) \tag{6}$$

$$\varepsilon_{jt}^i \sim N\left(0, \upsilon_j^2\right) \tag{7}$$

Note that the variance of the signal noise, υ_j^2, is variety-specific but time invariant. The advantage of this Gaussian framework is that the posterior distribution also has a normal distribution, $\theta_j^i \sim N\left(\overline{\theta}_{jt}^i, \sigma_{jt}^2\right)$. Furthermore, the mean $\overline{\theta}_{jt}^i \equiv E_t\left[\theta_j^i\right]$ and perception variance $\sigma_{jt}^2 \equiv Var_t\left[\theta_j^i\right] = Var_t\left[\theta_j\right]$, conditional on the information available after the signal at time t is revealed, satisfy the following updating equations (DeGroot, 2005):

$$\overline{\theta}_{jt}^i = \frac{\upsilon_j^2}{\sigma_{jt-1}^2 + \upsilon_j^2}\overline{\theta}_{jt-1}^i + \frac{\sigma_{jt-1}^2}{\sigma_{jt-1}^2 + \upsilon_j^2} s_{jt}^i \quad , \quad t = 1, 2, \ldots \tag{8}$$

$$\frac{1}{\sigma_{jt}^2} = \frac{1}{\sigma_{jt-1}^2} + \frac{1}{\upsilon_j^2} \quad , \quad t = 1, 2, \ldots \tag{9}$$

The Gaussian learning model, therefore, entails a simple elegant updating process: Both updating equations are linear, with the variance-updating equation being linear in the reciprocal of the variance, often called "precision" in this setting. Using Equation (9), the updating equation for the conditional variance can be alternatively stated as:

$$\sigma_{jt}^2 = \frac{\upsilon_j^2 \sigma_{j0}^2}{\upsilon_j^2 + t\sigma_{j0}^2} \tag{10}$$

From (9) and (10) we can derive the first important implication of this Bayesian learning model:

> **Result 1** As more information is acquired, the variance of the belief about the true agronomic optimal density θ_j^i decreases (i.e., $\sigma_{jt}^2 < \sigma_{jt-1}^2$), and as the number of signals goes to infinity, this perception variance converges to zero.

Next, consider the evolution of the conditional mean of the posterior distribution. Equation (8) conveys that the mean of the posterior distribution is a convex combination of the prior mean and the signal, and that the last signal becomes less and less important as t increases. Alternatively, by iterating the updating process in (8) and (9), the posterior mean can be expressed in terms of the prior and the sequence of innovation signals:

$$\overline{\theta}_{jt}^i = \frac{\sigma_{jt}^2}{\upsilon_j^2}\overline{\theta}_{j0}^i + \frac{\sigma_{jt}^2}{\upsilon_j^2}\left(\sum\nolimits_{k=1}^{t} s_{jk}^i\right), \quad t = 1, 2, \ldots \tag{11}$$

Recalling that $s_{jt}^i = \theta_j^i + \varepsilon_{jt}^i$, the posterior mean can be expressed as

$$\overline{\theta}_{jt}^i = \frac{t\sigma_{jt}^2}{\upsilon_j^2}\left(\theta_j^i + \frac{\overline{\theta}_{j0}^i}{t} + \frac{\sum_{k=1}^{t}\varepsilon_{jk}^i}{t}\right), \quad t = 1, 2, \ldots \tag{12}$$

From (10), $t\sigma_{jt}^2/\upsilon_j^2 = t\sigma_{j0}^2/\left(\upsilon_j^2 + t\sigma_{j0}^2\right) \to 1$ as $t \to \infty$. Therefore, it follows that $\overline{\theta}_{jt}^i \to \theta_j^i$ as t grows indefinitely. Hence:

> **Result 2** Information is useful: As the number of information signals increases, the mean of the posterior distribution converges to the truth, $\overline{\theta}_{jt}^i \to \theta_j^i$.

Note, however, that because the prior $\overline{\theta}_{j0}^i$ is unconstrained, the posterior $\overline{\theta}_{jt}^i$ can be either above or below the true value θ_j^i as it approaches it.

3.2 | Empirical specification

In the foregoing we have established that the optimal planting density is monotonically related to the prior belief about the agronomic optimal density. This belief is updated through learning as new information is acquired. The information signals of the Bayesian framework, in the empirical context of interest, can take several forms. Seller-recommended seed densities at the time of a new variety's introduction are inevitably coarse, being based on only limited trials in locations that may not represent the growing conditions faced by most farmers. This initial information set is supplemented by farmers' own experience: Upon using a given variety, yield realizations will provide an informative (however noisy) signal. Growers may also learn from the experience of other farmers with the same variety, or from their own experience with other, related, varieties—these are instances of two variants of Bayesian learning, "observational learning" and "correlated learning," respectively (Ching et al., 2013). Sharing of relevant information may also be mediated by the efforts of seed dealers, who are in a position to consolidate and disseminate signals of individual farmers, as well as the work of extension services. The latter may run field trials as well, to investigate the yield effects of

alternative seeding rates, something seed producers also continue to do in the early years of a new variety's commercialization.

Unfortunately, we do not have suitable data to disentangle the roles that such varied sources of information may have on the learning process. As the explicit Bayesian framework articulated in the foregoing makes clear, however, the number of information signals that is available prior to making a decision is crucial. We capture this effect by a variety-specific "age" variable: that is, the number of years a variety has been commercially available. One additional implication of Result 2 that we explore in the empirical analysis below specifically relates to the presumption that farmers learn from their own planting experience. Large farms—because they plant more fields and are likely to get more information signals in any one production cycle—may exhibit different rates of learning.

Factors other than learning signals, of course, may affect the observed chosen planting density, including the seed price, the expected output price, as well as other farm specific factors. Hence, to empirically investigate the learning processes as revealed through observed planting densities, we estimate the following fixed effects regression equation:

$$z_{ijt} = \beta Age_{jt} + \phi R_{ijt} + \alpha_i + \lambda_j + u_{ijt}, \tag{13}$$

where z_{ijt} is the observed planting rate (kernels/acre) for variety j by farmer i in year t, and $R_{ijt} \equiv w_{ijt}/p_t$ is the ratio of the observed seed purchase price w_{ijt} (\$/80,000 kernels, and again for variety j planted by farmer i in year t) to the expected maize price p_t, measured by the harvest-time maize contract futures price quoted at planting time (\$/bu). The key variable of interest is Age_{jt}, which measures the number of years variety j has been commercially available as of year t. For example, for a variety first introduced in year 2003, this variable takes value 1 in 2003, value 2 in year 2004, and so on.

The term λ_j is a variety fixed effect. This term is essential as it captures the fact that, as discussed earlier, modern varieties are associated with higher planting densities, ceteris paribus. Thus, having introduced variety fixed effects, the results can then be interpreted as pertaining to a fixed technology. The term α_i is a farmer fixed effect, which is motivated by Equation (4), wherein it was recognized that the optimal planting density, for a given variety, may vary across farms (which are heterogeneous with respect to land and climatic conditions). Inclusion of the farmer fixed effect in the empirical model is possible because, as noted earlier, in the data many farmers are observed to plant more than one variety on a given farm.[2] These fixed effects are also essential for estimating and interpreting the Age_{jt} variable coefficient. Thus, the coefficient β captures whether and how quickly farmers update their priors about the optimal rate, on average (i.e., across all varieties). As a consequence of Result 2 derived earlier, the Bayesian learning model does not constrain the sign of this coefficient. This sign depends on the nature of farmers' prior beliefs, specifically on whether farmers' initial priors for the AOPD were below or above their true values, on average. For example, if farmers' priors were unbiased, such that $\theta_{j0} \cong \theta_j$ on average, then one should expect β to be zero. Note also that because β is not variety specific, we are implicitly assuming that the rate of learning is the same across varieties. Alternatively, β can be interpreted as a mixture of different underlying variety-specific learning rates. In a set of robustness checks, we relax this constraint and allow β to be heterogeneous across varieties.

Recall that Result 1 establishes that the variance of the posterior of the AOPD decreases as information accumulates. As argued in the foregoing, this also means that the variance of observed variety-specific planting densities should decrease as information accumulates. This implies that the error term in Equation (5) is heteroskedastic in the variety's commercial age. In particular, the variance of this error term should decrease when information accumulates. Therefore, we also estimate the following regression equation:

[2]In other words, because a farmer's various varieties are likely to be planted on land that share common soil and climatic conditions, the farmer fixed effect can control for some of the heterogeneity across farms seeded by the same variety. An alternative control that we report in the empirical analysis is to replace farmer fixed effects by regional fixed effects (specifically, crop reporting district [CRD] fixed effects).

$$ssr_{ijt} = \gamma Age_{jt} + \tau_i + \kappa_j + e_{ijt}, \tag{14}$$

where the dependent variable pertains to the standardized squared residuals from the estimated planting density equation, that is $ssr_{ijt} \equiv \left(z_{ijt} - \hat{z}_{ijt}\right)^2 / \hat{\sigma}_u^2$, where $\hat{\sigma}_u^2$ is the estimated variance of the residuals of Equation (13), whereas τ_i and κ_j are farmer and variety fixed effects, respectively. Result 1 implies that γ is negative if learning occurred.

4 | DATA

The empirical models are estimated with the maize TraiTrak® dataset, collected by the company Kynetec USA, Inc. These data consist of farm-level seed sales data spanning the period 1995–2016. The data were assembled from annual surveys of randomly sampled US farmers, with the samples designed to be representative at the crop reporting district (CRD) level. Over the period of analysis, the dataset contains surveys from an average of 4733 maize farmers per year. For each sampled farmer, we observe the planting rate, identities of varieties planted, and price paid. On average, each farmer purchases 4.23 varieties per year. Some farmers are surveyed over multiple years, which permits the inclusion of farmer fixed effects.

The original dataset provides 442,803 observations on chosen planting densities, where an observation is a unique combination of the year, farmer, and planted hybrid. In order to estimate the model, we trim the dataset in two ways. First, in some cases a farmer did not report the identity of the variety planted. Because we cannot include variety fixed effects for these observations, we drop them from the dataset, reducing the sample to 403,327 observations. Second, we remove the very small number of cases where a value of zero was recorded for the planting rate. Our finalized dataset consists of 403,261 observations across the 1995–2016 period. In some specifications, we also include variables for the share of acres (at the CRD level) cultivated with reduced-till technology and the number of neonicotinoid acre treatments (at the CRD level).[3] These variables had limited availability during the 1998–2016 timeframe, and thus the specifications that include these variables have a further reduced sample size of 360,528 and 314,393, respectively.

Table 1 contains statistics that summarize the structure of the final sample consisting of 403,261 observations. On average, 4412 farms were sampled per year. Although this is not a balanced sample, farms are typically sampled more than once. On average, a given farm was sampled 2.77 years over the interval, but some farms were sampled all 22 years (just three). Each farm purchased 4.23 distinct varieties per year, on average, and each farm planted the same variety 1.22 times across the years they were sampled. This last statistic, however, also reflects how many years a farm was sampled. For example, if we only observe a farm in two different years, we can only observe them planting a given variety at most twice. This observation is another reason in support of using a commercial age variable that is not specific to each farm. Given that most farms are observed a fraction of the years in the sample, variables that track how many times a farmer planted a variety would significantly underestimate the number of times that farm had actually purchased a particular variety.

Table 2 reports summary statistics for each of the variables used in the finalized sample. The mean overall planting rate was 29,532 seeds per acre. The table also reports planting rates disaggregated by the central maize belt (CCB) and the non-CCB, where the CCB includes IA, IL, IN, and the southern crop reporting districts in MN and WI. The average planting rate in the CCB was 31,047, significantly higher than the non-CCB rate of 28,091. On average, the commercial age for a variety was 2.77 years, although some varieties were actually observed for the maximum possible value of 22 years. A related variable is the life cycle of a variety, which measures how many years a variety was commercially available. The average life cycle for a variety was about 6.3 years, a

[3]Neonicotinoid acre treatments are the ratio of the number of acres treated with a neonicotinoid insecticide to the number of planted acres.

TABLE 1 Dataset summary

Variable	Mean	*SD*	Min	Max
Farms sampled per year	4412	564	3092	5194
No. years sampled per farm	2.77	2.78	1	22
Unique varieties per year	3715	683.3	2769	5073
No. purchases/farm/year	4.23	3.05	1	41
No. times given variety purchased	1.22	0.58	1	12

Note: there are 34,363 unique farms sampled in the data.

TABLE 2 Summary statistics for model variables

Variable	*N*	Mean	*SD*	Min	Max
Planting rate					
Overall	403,261	29,532	4509.4	8000	57,143
Non-CCB	206,660	28,092	4899.9	8000	57,143
CCB[a]	196,601	31,047	3463.9	10,000	53,333
Commercial age	403,261	2.774	1.826	1	22
Life cycle	403,261	6.289	3.790	1	22
Seed price	403,261	155.06	80.326	0	420
Maize futures[b]	403,261	3.737	1.233	2.324	5.894
Price ratio	403,261	40.624	14.694	0	113.5
Reduced-till share	360,528	0.587	0.200	0	1
Neonic acre treatments	314,393	0.551	0.427	0	2.000

[a]IA, IL, IN, and the southern crop reporting districts in MN and WI.
[b]Annual data (there are 22 unique values).

relatively short time span. Seed prices averaged $155.06 per 80,000 kernels, and the maize futures price averaged $3.74 per bushel; the mean ratio of these prices was 40.62.

5 | RESULTS

Table 3 contains regression results for the baseline model. We report results for four specifications, each differing by the type of fixed effects included. The importance of fixed effects is demonstrated most starkly by comparing Column 1, which contains no fixed effects, with Column 2, which contains variety fixed effects. In Column 1, the coefficient on the price ratio variable is positive and significant, contrary to expectations, and the coefficient on the commercial length variable is negative and significant. Both coefficients flip signs upon introducing variety fixed effects. Intuitively, the estimated coefficients in Column 1 are, in part, based on comparisons of planting rates for newer varieties, that is, those with short commercial life spans, to planting rates for older varieties that are still on the market. By contrast, the fixed effects estimator in Column 2 is based on within-variety variation.

The final two columns introduce regional or farmer fixed effects. In general, they confirm the presence of significant unobserved factors that are both correlated with seed age and planting rates. Column 3, for example, adds CRD fixed effects in addition to variety fixed effects and results in a larger estimate for the seed age variable (about 126 kernels compared to 92.6 kernels). This suggests that newer varieties are first introduced in higher planting rate regions (such as the CCB), and then

diffuse to lower planting rate regions. Column 4 replaces CRD fixed effects with farmer fixed effects, which has the effect of increasing the coefficient on seed age even further to a point estimate of about 222 kernels per year. This suggests that early adopters of new varieties tend to plant at significantly higher rates compared to late adopters.

It is also of some interest to note that the responsiveness of farmers' planting rates to market prices (the ratio of seed to expected output price) is extremely small. Recall that the sample average of w/p is approximately 40.6. Hence, a 10% increase in this price ratio, given the estimated coefficient in the last column of Table 3, would reduce planting rates by a mere 8 kernels per acre (less than 0.03% of the average planting rate)!

Table 4 presents results for the squared residual regressions. Specifically, using the coefficient estimates from specification (4) in Table 3, our preferred specification, we compute the predicted standardized squared residuals for each observation and then regress these on the commercial age variable. Having established the importance of including fixed effects, here we report estimates from two specifications, one with regional and the other with farmer fixed effects. Overall, we find significant evidence of decreasing planting rate variance, and therefore of learning, over time. The fourth specification, which includes variety and farmer fixed effects, indicates that each additional commercial year is associated with a reduction in planting rate variance by roughly 4% per year.

5.1 | Robustness checks and alternative explanations for inertia

Overall, the results from Tables 2 and 3 confirm two facts: (i) the variance in planting rates, conditional on variety and farmer fixed effects, decreases significantly over time; and (ii) variety-conditioned planting rates increase over time. This latter result is particularly noteworthy, as it suggests there is a persistent bias in farmers' choices: more often than not, they *underinvest* in density at the onset of a hybrid's commercial life. As previously noted, however, additional explanations may contribute to this result. Thus, in this section we consider some of these alternative explanations and factors while also attending to some other potential limitations of our baseline framework.

5.1.1 | Remove early hybrids

For hybrids that appear early in the sample, the commercial age variable may suffer from truncation. For example, the true commercial age of varieties purchased in 1995 is unknown. Thus, the imputed

TABLE 3 Maize planting rate regressions

	(1)	(2)	(3)	(4)
Commercial age	−71.42**	92.55**	126.15**	222.03**
	(6.45)	(7.02)	(5.79)	(4.74)
Price ratio (w/p)	90.65**	0.14	−3.92**	−1.43*
	(1.13)	(1.26)	(1.03)	(0.75)
Constant	26,047.9**	29,291.8**	29,362.3**	29,013.4**
	(59.67)	(58.71)	(47.49)	(32.25)
Fixed effects	None	Variety	Variety, CRD	Variety, farmer
Observations	403,261	403,261	403,261	403,261
R^2	0.089	0.340	0.509	0.783

Note: For all regressions, the dependent variable is the planting rate. Models in each column differ by the included fixed effects. Standard errors, clustered at the farm-level, are in parentheses.
$^*p < 0.10$. $^{**}p < 0.01$.

TABLE 4 Standardized squared residuals regressions

	(i)	(ii)
Commercial age	−0.031*	−0.041*
	(0.005)	(0.006)
Constant	1.086*	1.116*
	(0.018)	(0.016)
Fixed effects	Variety, CRD	Variety, farmer
Observations	386,872	386,873
R^2	0.095	0.282

Note: The dependent variables are the standardized square residuals from Model (4) in Table 3. The models in each column differ by the included fixed effects. Standard errors, clustered at the farm level, are in parentheses. Observation numbers are smaller than those in Table 3 because observations with perfect fit (due to the fixed effects) are omitted.
*$p < 0.01$.

commercial age for many varieties will be too low and their relative age will be incorrect. To assess the importance of this issue, we remove all varieties observed in the first five years of the sample (1995–1999). Given the short commercial life of most hybrids, discussed earlier, it is extremely unlikely that varieties first observed in 2000 or later had actually been available prior to 1995. Column 1 of Table 5 reports the results for this reduced sample with a more accurate measurement of the age variable. We find that the coefficient of this variable is essentially unchanged. Thus, in the remaining columns we include varieties from the entire sample.

5.1.2 | Additional controls

The second and third columns of Table 5 add control variables for the share of acres with reduced tillage (at the CRD level) and the number of neonicotinoid acre treatments (at the CRD level). Both practices became more prevalent during the study period and are potentially related to the seeding rate (Perry et al., 2016; Perry & Moschini, 2020). For example, some farmers report that neonicotinoid treated seed have a higher germination rate, which permits a lower seeding rate for a target plant population. Overall, the coefficient on the commercial age variable is robust and not significantly affected by the addition of these variables.

5.1.3 | Regional and individual differences in updating

The baseline estimates do not permit heterogeneity in the response of planting rates to the commercial age variable. Yet, a large literature points to characteristics such as farm size, productivity, and demographic variables that are correlated with farmer decision making. For example, some recent evidence indicates that smaller farms are slower to update their pesticide application behavior in response to the availability of new products (Perry et al., 2019). In addition, the summary statistics in Table 2 indicate that the more productive regions—the CCB—are associated with higher planting densities. It may also be the case that these regions increase planting rates at different rates, perhaps due to larger increases in productivity over time. To assess some of these possibilities, Columns 4–6 present estimation results that allow the commercial age variable to be heterogeneous. Columns 4 and 5 contain separate estimation results for the CCB and the non-CCB, respectively. The commercial age coefficient is almost 40% larger in the CCB compared to the non-CCB. Column 6 presents results for a model in which the commercial age variable is allowed to differ over three farm size ranges: (i) planted maize acres were less than 100; (ii) maize acres were between 100 and 1000; and (iii) maize acres were greater than 1000 acres. The estimation results show that larger operations

TABLE 5 Additional maize planting rate regressions

	(1) Remove early varieties[a]	(2) Add neonic variable	(3) Add till variable	(4) CCB	(5) Non-CCB	(6) Farm size
Commercial age	216.03***	227.86***	214.89***	260.72***	188.18***	
	(5.60)	(8.15)	(4.92)	(5.99)	(7.10)	
Commercial age × acres < 100						246.19***
						(9.95)
Commercial age × 100 < acres < 1000						221.12***
						(5.06)
Commercial age × 1000 < acres						208.05***
						(8.57)
Price Ratio (w/p)	−1.08	−0.45	−1.52**	−1.66*	−1.57	−1.40*
	(0.76)	(0.85)	(0.75)	(0.95)	(1.18)	(0.75)
Neonic acre treatments		−75.21				
		(79.23)				
Reduced till share			−381.18***			
			(84.61)			
Constant	29,958***	29,066***	29,571***	30,462***	27,628***	29,013***
	(35.74)	(47.23)	(61.29)	(42.39)	(49.08)	(32.25)
Observations	284,578	314,393	360,528	196,601	206,660	403,261
R^2	0.819	0.788	0.796	0.745	0.773	0.783

Note: For all regressions, the dependent variable is the planting rate. All models include farmer and variety fixed effects. Standard errors, clustered at the farm-level, are in parentheses.
[a]This specification removes data for varieties first observed during the period 1995–1999.
*$p < 0.10$. **$p < 0.05$. ***$p < 0.01$.

TABLE 6 Variety-specific regressions for the top 15 hybrids

Hybrid	Seed age coefficient	Standard error	Observations	R-squared
3394	180.002***	60.769	2537	0.74
33A14	221.192*	130.965	1205	0.681
33B51	176.472*	101.471	1026	0.73
33G26	203.792**	84.644	1263	0.737
33P67	252.624***	73.323	1206	0.695
3489	311.518***	83.927	1477	0.685
34B23	76.153	89.923	1221	0.685
36B08	240.556***	70.614	1150	0.656
3730	355.255***	92.247	1319	0.688
3751	202.252**	86.483	1402	0.651
DKC48-12	2.596	80.271	1021	0.876
DKC52-59	269.870***	48.386	1684	0.89
DKC52-62	289.053***	50.818	1275	0.866
DKC61-69	374.886***	70.636	1594	0.785
DKC63-42	323.966***	72.495	1294	0.866

Note: All models include the price ratio variable and farmer fixed effects. The standard errors are clustered at the farm-level.
*$p < 0.10$. **$p < 0.05$. ***$p < 0.01$.

increase densities at a lower rate. For example, a farm with more than 1000 maize acres increases seeding rates, on average, at a rate of about 208 seeds per year, nearly 40 kernels per year less than farms with less than 100 maize acres. An additional fact, not reported in Table 5, is that larger farms tend to *start* at higher planting rates. In the context of the learning model, these two facts taken together suggest that larger farms are, on average, less biased in their priors for the AOPD.

5.1.4 | Variety-specific regressions

The learning model presented above permitted learning to differ across seed varieties. However, for simplicity, the baseline empirical framework set the commercial age coefficient to be the same across varieties. To assess the degree of heterogeneity in the commercial age coefficient across varieties, Table 6 contains the commercial age coefficients for models estimated separately for the 15 most popular hybrids in our sample. These hybrids correspond to the same hybrids presented in Figure 2 above. In contrast to the trends presented in that figure, these trends are estimated in models that also include farmer fixed effects and the seed-output price ratio. Of the 15 seed age coefficients, 13 are positive and statistically significant. The other two cases are also positive but statistically insignificant. The values range from about 2.6 kernels per year for DeKalb hybrid DKC48-12 to a high of nearly 375 kernels per year for the DeKalb hybrid DKC61-69. Overall, the variety-specific regressions show some heterogeneity in the estimated responses, but the direction of updating is invariably positive and the magnitude remarkably consistent with the aggregate results of Table 3.[4]

[4]Based on a suggestion by the editor, we also estimated regressions with a cumulative individual purchase variable, which tracks the number times a variety was purchased by a farm up to the current period. The motivation for including this additional variable is to try to disentangle the role of individual experience from market experience as drivers of learning about planting rates. Overall, we do not find evidence that individual experience is a driver of the upward trend in planting rates. Individual experience does, however, appear to reduce the perception error variance. We note, in any event, that the data at hand pose limitations to this line of inquiry: There are simply not many instances in the data in which a farmer purchases a particular variety more than once. For further details, see the Online Supplementary Appendix.

5.2 | Risk aversion and nonlinearities

Our baseline analysis assumes that farmers are risk neutral vis-à-vis their seeding rate choices. A relevant question is how the presumption of risk aversion would affect the conclusions to be drawn from the empirical analysis. Recognizing that the true agronomic optimal seeding rate $\widetilde{\theta}$ is uncertain, from the stylized model presented earlier, per-acre random profit can be expressed as:

$$\widetilde{\pi} = pb\widetilde{\theta}z - 0.5pbz^2 - wz \tag{15}$$

Note that the random variable $\widetilde{\theta}$ enters linearly in per-acre profits. Indeed, the structure of this equation is isomorphic to that of a competitive firm facing price uncertainty with a convex (quadratic) cost function. Standard analysis (e.g., Moschini & Hennessy, 2001) then would show that a risk-averse farmer who maximizes the expected utility of this profit, $E[u(\widetilde{\pi})]$, would choose a seeding rate z_t^{A*} that is lower than the seeding rate z_t^{N*} chosen by a risk-neutral agent. Furthermore, it would follow that z_t^{A*} increases as uncertainty about $\widetilde{\theta}$ is resolved because of learning.

This framework permits some conclusions regarding how risk aversion affects our baseline analysis. Learning implies that uncertainty about the agronomic optimal density $\widetilde{\theta}$ is reduced as more information is collected. This means that σ_t^2 decreases, such that the risk-aversion effect, per se, will tend to be associated with an increasing planting density over time. Of course, learning also affects the mean $\overline{\theta}_t$. Thus, for any given prior we have two general cases:

i. When $\overline{\theta}_0 < \theta$, such that the prior mean is below the true value, the mean will be revised upward as more information is collected. Learning also reduces the posterior variance of the parameter, and because of risk aversion, this effect tends to increase the planting rate. Hence, when the prior is biased downward, the mean effect is reinforced by the risk-aversion effect, and the overall effect is unambiguous: planting rates increase as a consequence of learning.
ii. When $\overline{\theta}_0 > \theta$, such that the prior mean is above the truth, the mean is revised downward as more information is collected. In this case, learning about the mean tends to drive down the planting rate, whereas the reduced variance due to learning would tend to increase the planting rate. Because the mean effect and the risk-aversion effect work in opposite directions, in this case the overall effect is ambiguous.

These observations have ancillary econometric implications: Identification of the risk aversion effect, separate from the mean effect, requires information and/or hypotheses about how priors are formed. As in related work, however, we do not attempt a formal model for the formation of prior beliefs. The difficulty of separating the risk-aversion effect from agents' priors has been recognized by previous work. Much of the literature on learning in agriculture, for example, Conley and Udry (2010), assumes that farmers are risk neutral.

The foregoing risk-aversion analysis is predicated on the quadratic representation of the yield response to planting density, which conveniently implies that the per-acre profit in (15) is linear in the parameter θ, the object of learning. Similar implications arise if, instead, we continue to presume risk neutrality but the object of learning enters the decision problem nonlinearly. To illustrate, suppose that, as in Gaspar et al. (2020), the per-acre yield function is written as

$$y = y_{\max} \times \left[1 - e^{-\theta z}\right] \tag{16}$$

where z continues to denote seed density, $y_{\max}$ is the asymptotic yield maximum, and $\theta > 0$ is a parameter that determines the responsiveness of yield to seeding rates. Recall that the appeal of the earlier quadratic yield representation is the presumption that the crowding of plants will eventually decrease per-acre output, which has a clear agronomic basis. The yield function in Equation (16), on

the other hand, does not admit a negative marginal product: $\partial y/\partial z = \theta(y_{\max} - y) > 0$. Still, this parameterization may be quite useful as a local approximation of the yield response to plant density.[5]

Suppose that, with the parameterization in (16), farmers learn about the parameter θ, and that (consistent with the Gaussian Bayesian model) this parameter is normally distributed, $\theta \sim N(\overline{\theta}, \sigma_\theta^2)$. Then the expected yield, for any given seeding density z, is

$$E\left[y\left(\widetilde{\theta}\middle|z\right)\right] = y_{\max}\left(1 - \int e^{-\theta z}\frac{1}{\sigma_\theta\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\theta-\overline{\theta}}{\sigma_\theta}\right)^2}d\theta\right) \tag{17}$$

The random per-acre profit here is $\widetilde{\pi} = py\left(\widetilde{\theta}\middle|z\right) - wz$. The integral in (17) admits a closed-form solution such that the per-acre expected profit can be expressed as:

$$E[\widetilde{\pi}] = p\, y_{\max}\left(1 - e^{-z\left(\overline{\theta} - \frac{1}{2}z\sigma_\theta^2\right)}\right) - wz \tag{18}$$

It can now be shown that the optimal seed density z^* that maximizes this per-acre profit is decreasing in the variance σ_θ^2 whenever it is increasing in the expected value $\overline{\theta}$.[6]

Thus, the presumption that farmers are risk averse with a quadratic yield function or that farmers are risk neutral but with an exponential yield function both lead to similar conclusions—the effect of learning, via the reduction in the variance of the parameter of interest, would tend to increase optimal planting rates. As a result, the empirical results discussed earlier are consistent with the presumption that farmers have correct (unbiased) priors, on average, but they are risk averse and/or the yield function is nonlinear in the parameter that is the object of learning.

Whereas we recognize that risk aversion and/or nonlinearities may play a role in this context, we do not believe the empirical results uncovered here are mostly due to such effects. In fact, the analysis of soybean planting rates provides insightful corroborating evidence.

5.3 | Soybean seeding rates

Soybeans constitute the second most important U.S. row crop, and most farmers in our sample plant soybeans in rotation with maize. Using Kynetec data on observed seeding rates for soybeans, we estimate the same set of baseline regressions that were estimated for maize. Table 7 contains the estimation results for soybeans. We report the results of both the planting rate equation and the standardized squared residuals equation. Having explained the need for variety fixed effects earlier, here we focus on the two specifications of the planting rate equation that include variety fixed effects (Columns (1) and (2)).[7] In contrast to maize, we find that within-variety soybean seeding rates are decreasing with learning, as captured by the "commercial age" variable. This result appears to conflict with that for maize but is actually consistent with the agronomic differences between the two crops. In fact, agronomical research has emphasized that optimal seeding rates for soybeans have *fallen* over time (de Bruin & Pedersen, 2008; Gaspar et al., 2020; Lee et al., 2008). Farmers' beliefs, as revealed by their observed choices, appear to be catching up with this agronomic evidence. That

[5]Experimental data reported in Coulter (2021), for example, show that yields in Minnesota are still not decreasing at 45,000 kernels/acre, although they typically flatten at about 33,000 kernels/acre.

[6]Just as in the case of certainty with the yield function (16), the optimal seed density z^* is increasing in the expected value of the θ parameter as long as the price ratio w/p is not too small.

[7]Soybean planting data are available to us for the period 1996–2016. To interpret the reported estimated coefficients, note that the mean value of the left-hand-side variable (planting rate) is 168,761 seeds/acre, the mean value of the price ratio is 3.82, and the mean value of the commercial age variable is 2.98.

TABLE 7 Soybean regression results

	Planting rates		Standardized squared residuals	
	(1)	(2)	(i)	(ii)
Commercial age	−818.99*	−1022.15*	−0.035*	−0.052*
	(57.71)	(55.32)	(0.005)	(0.006)
Price ratio (w/p)	−1268.54*	−1069.35*		
	(133.37)	(114.13)		
Constant	175,914.30*	175,718.85*	1.107*	1.159*
	(574.92)	(454.64)	(0.020)	(0.018)
Fixed effects	Variety, CRD	Variety, farmer	Variety, CRD	Variety, farmer
Observations	187,776	187,776	175,832	175,832
R^2	0.334	0.683	0.104	0.313

Note: For models in columns 1 and 2, the dependent variable is the planting rate. For models in columns (i) and (ii), the dependent variable is the standardized squared residuals of the corresponding planting rate equation. The number of observations is smaller for the standardized squared residuals regressions because observations with perfect fit (due to the fixed effects) are omitted. Standard errors, clustered at the farm-level, are in parentheses.
*$p < 0.01$.

learning is actually taking place here as well, similar to maize, is confirmed by the standardized squared residuals equation results (columns (i) and (ii) in Table 7). Just as for maize, we find that accumulated experience with a given variety reduces the dispersion of the observed planting rates.

The foregoing results for soybeans, furthermore, weaken the case for risk aversion, and/or nonlinearities, as the main explanation for the observed pattern in maize. The choice of soybean seeding rates is structurally similar to that of maize: if risk aversion and/or nonlinearities were the primary determinant of variety-specific temporal adjustment in seeding rates, then, given unbiased priors (on average), one should also expect soybean seeding rates to increase over time, ceteris paribus. But in fact, it seems that learning is decreasing soybean seeding rates, just as learning appears to lead to increasing maize seeding rates. In both cases, the estimated adjustments are in the direction suggested by the prevailing agronomic wisdom. Thus, it appears that for both crops we have uncovered evidence of "inertia" in response to information, a phenomenon that has long been observed in laboratory-based studies (Benjamin, 2019; Henckel et al., 2021; Phillips & Edwards, 1966) and also in a variety of real-world circumstances (Handel & Schwartzstein, 2018; Perry et al., 2019). The phenomenon may arise for many reasons, including selectivity when paying attention to data such that biased beliefs and forecasts result (Schwartzstein, 2014). A literature has emerged showing that experience with context tends to eliminate non-standard economic responses and investors in some instances (List, 2003) but accentuate them in others (Haigh & List, 2005). However, unlike most of the extant literature, our context regards important real-world business decisions and it involves producers' rather than consumers' or investors' choices. We find, nonetheless, a clear instance of inertia—farmers demonstrate a persistent disposition toward choosing planting rates that may have been appropriate with past varieties but are no longer optimal for new varieties.

5.4 | Counterfactuals

Using the estimated model, we assess two counterfactual scenarios: (i) the case in which the seed age variable is set to zero; and, (ii) the counterfactual in which 6.3 years (i.e., the average commercial life cycle of a variety) are added to the seed age variable. The first scenario evaluates the consequences of no learning, whereas the second scenario assesses the significance of endowing farmers with more information from the beginning. The latter scenario also informs on the implications of chronic

underinvestment in seeding rates by farmers. For comparison purposes, we first generate a baseline by predicting planting rates using the estimated coefficients from column 4 of Table 3:

$$\hat{z}_{ijt} = \hat{\beta} Age_{jt} + \hat{\phi} R_{ijt} + \hat{\alpha}_i + \hat{\lambda}_j. \tag{19}$$

These predicted seeding rates can be compared with two counterfactual predictions. One is the seeding rates $\widetilde{z}_{ijt}$ that would be expected if no learning took place, which are obtained by setting the commercial age variable to zero, that is:

$$\widetilde{z}_{ijt} = \hat{\phi} R_{ijt} + \hat{\alpha}_i + \hat{\lambda}_j. \tag{20}$$

Alternatively, we can consider the seeding rates $\breve{z}_{ijt}$ that would arise if farmers were endowed with considerably more information from the beginning, that is, at the time of the variety's initial release. Specifically, suppose that the information equivalent to 6.3 years of experience with any given variety—the average life cycle of a commercial hybrid variety in our sample—is a good approximation to what is likely knowable for a mature variety. Then, the counterfactual seeding rates associated with this augmented information is estimated as:

$$\breve{z}_{ijt} = 6.3\hat{\beta} + \hat{\phi} R_{ijt} + \hat{\alpha}_i + \hat{\lambda}_j. \tag{21}$$

Using these predicted seeding rates, we next calculate the national average for each year in the sample. The annual predictions associated with each counterfactual, as well as the baseline annual predictions, are reported in Figure 3. Comparing the baseline to the counterfactual without learning, there is a clear downward shift in planting rates by 625 kernels per year, on average, equivalent to about 2.1% of mean planting rates. Conversely, with the information level equivalent to 6.3 years of experience with a variety, seeding rates would shift up by about 774 kernels per year, on average, equivalent to about 2.6% of observed planting rates.

The planting rate counterfactuals have implications for expected yields. Whereas we do not observe farm-level yields, we can link seeding rate predictions to yields by using yield-density relations from field trial data. In particular, Assefa et al. (2016) estimate yield response functions using density trial data from DuPont Pioneer. Their data span the period 2000–2014 and consist of 124,374 observations for Pioneer hybrids under different planting densities. They find that a quadratic model fits the data best. To be consistent with such experimental data, here we restrict our attention only to Pioneer hybrids and focus on hybrids planted in Iowa and Illinois. These are the two most significant maize growing regions in the United States, and the goal of this exercise is only to obtain a sense of the order of magnitude associated with the planting rate counterfactuals. In Assefa et al. (2016), the relevant estimated quadratic equation for the latitudes corresponding to Iowa and Illinois is:

$$\hat{y}_{ijt} = 3.29 + 0.18\hat{z}_{ijt} - 0.001\hat{z}^2_{ijt}. \tag{22}$$

Using this equation, along with the predicted values of z_{it} for Iowa and Illinois based on the CCB coefficient estimates from Column 4 of Table 5, provides the baseline yield predictions. We then generate new predictions for z_{it} using Equations (20) and (21), and insert these values into Equation (22). The difference in yields between these scenarios constitutes an estimate of the impact of learning and additional information, respectively.

For the scenario without learning, the average reduction in yields is about 1 bu/acre per year, or 0.6% of mean yields in IA and IL. There is also considerable heterogeneity in these effects. For example, the 5th percentile effect is about 0.2 bu/acre, whereas the 95th percentile effect is about 2.7 bu/acre. To provide additional context, at a maize price of $4/bu, an impact of 1 bu/acre translates to $4/acre in additional revenue. Total maize acreage in Iowa and Illinois in 2019 was around

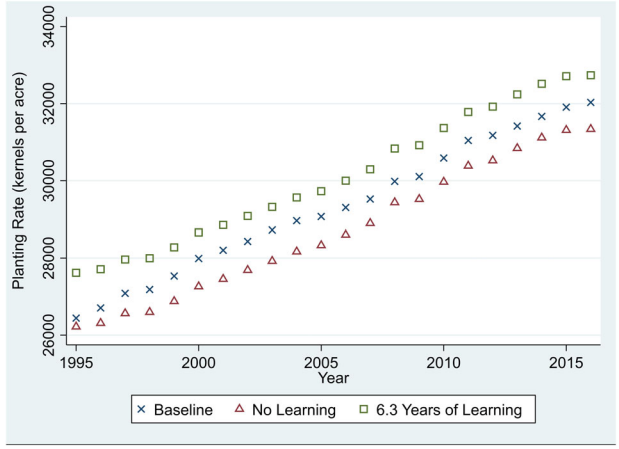

FIGURE 3 Predicted counterfactual seeding rates. Notes: This figure shows predicted planting rates under three different scenarios, based on Equations (19)–(21) using the estimated parameters from column 4 of Table 3. The "Baseline" predicted rates are generated with the observed commercial Age variable. The "No Learning" predicted rates are generated with the commercial Age variable set to zero. The "6.3 years of Learning" rates are predicted with the variable Age = 6.3 (the average commercial life cycle of a variety in the sample)

25 million acres. Thus, this effect results in estimated additional $100 million in farmers' revenue. Given that average yields in these two states was about 190 bu/acre in 2019, this is equivalent to 0.5% increase in revenue per farm. The counterfactual with information equivalent to 6.3 years of learning, available immediately when the variety is first released, generates effects of a similar magnitude but in the opposite direction: on average, yields would increase by 0.85 bu/acre, or about 0.5% of mean yields. Of course, these are gross effects and not net of possible additional costs, and these calculations are simply meant to provide a yardstick. Two main points seem worth emphasizing, however. First, the revenue impact of learning about seed density is not insignificant—the difference between the best information scenario considered (the "6.3 years of learning") and no learning is about 1% of farmers' maize revenue. Second, the amount of learning in the market, as displayed by the data, is not trivial—about half of that associated with the best information scenario considered.

6 | CONCLUSION

There is continuing interest in obtaining a deeper understanding of the root causes of the exceptional technological progress that has characterized modern agriculture (Olmstead & Rhode, 2008; Wright, 2012). Our paper contributes unique empirical evidence concerning an important link between improved maize varieties and increased yields: higher planting rates. Based on an extensive dataset of observed seed planting densities by a representative sample of U.S. maize farmers, covering more than 400,000 purchases over 22 years, we have established some clear empirical findings. First, farmers' planting density choices display a pattern unambiguously consistent with Bayesian learning—for a given variety, as more information about the variety accumulates, the variance of chosen planting rates decreases. A second strong empirical finding is that a given variety tends to be planted at higher densities as more information is accumulated. This finding is prima facie consistent with learning—if everything relevant to the optimal planting density were known when the variety is first commercialized there would be no reason to expect planting densities to change over time (again, for a given variety). On closer inspection, however, this finding is revealing because the result here is about the average tendency across all varieties, conditional on variety fixed effects. Hence,

our results suggest that, on average, farmers' initial priors about varieties' optimal planting densities are systematically below the truth.

Several considerations may be germane in interpreting these empirical results. First, learning is necessarily slow in our setting because a signal about a variety's optimal planting density takes a full production cycle (1 year) to materialize. Second, the degree to which farmers learn from the experience of others about a given variety is limited by heterogeneity across space and time (e.g., differing latitude, soil conditions, and weather). Learning from experimental results produced by extension services and seed companies is also limited by the fact that such trials only typically consider a select number of varieties, as well as by a coarse grid of possible planting densities in these trials. One could add, furthermore, that farmers' skepticism toward the recommendations of seed companies may be due to imperfectly aligned incentives (farmers can lose by under seeding or by over seeding, whereas seed sellers' profit only increases with seed density). Indeed, Hennessy et al. (2021) find that maize farmers place the most weight on their own experience with a variety in determining the best planting rate, and the recent rise in distributed farmer networks (such as the Farm Business Network) have been driven in part by farmers' need for information beyond what is provided by seed sellers and their own experience.

Whereas we have presented compelling evidence of learning behavior by US maize farmers, the evidence we have uncovered shows that this learning process is slow and affected by inertia. These findings support the notion that learning is inherently difficult in complex environments, and several elements of complexity matter in our context. As noted earlier, the life cycle of commercial maize varieties is short, and farmers have to form priors for continuously introduced new varieties. Furthermore, in a technologically dynamic industry, farmers often learn about more than just the optimal planting density. This is particularly true of the U.S. maize industry during the period we examine, which has seen the introduction and widespread adoption of GE seed varieties (Ciliberto et al., 2019). Over time, the progressive embedding of commercialized varieties with multiple GE traits (glyphosate tolerance and insect resistance) has led to an even more complex decision environment. It is perhaps unsurprising, therefore, that in learning about the optimal density, farmers may have often relied on their own past experience. Consequently, as suggested by the empirical results in this paper, farmers' planting choices display a degree of inertia vis-à-vis the optimal planting density.

Maintaining crop productivity remains a major tool to address the challenges of feeding a growing world population in the midst of climate change. U.S. maize yields have attracted considerable interest in this context, and there is a general recognition that continuing innovation efforts are essential (Lee et al., 2021; Ortiz-Bobea & Tack, 2018). An important implication of our findings is that there may be unrealized returns from efforts intended to assist in optimizing planting rates for new maize hybrids. Additional variety trials, however conducted, could help shift up the yield curve. More generally, whereas research into the development of new agricultural technologies is very important for continued productivity growth, our results emphasize the belief that attention to the implications of the increasing complexity accompanying such technologies should not be neglected. Learning processes associated with technology adoption and diffusion remain of paramount importance, and they are not confined solely to (widely studied) developing countries' contexts. How learning can be best assisted is a matter for debate (Norton & Alwang, 2020), but new opportunities may be available with a more extensive deployment of information technologies and precision agriculture practices. In particular, fostering peer-to-peer interactions may be well-suited to learning about the optimal seeding rate choice because much of the information resides with farmers themselves. The recent growth in precision agriculture has led to the emergence of large distributed databases on farming outcomes. Thus, one potential policy that could encourage the diffusion of information would be the allocation of public funds toward incentivizing enrollment in peer-to-peer databases that track yield outcomes associated with different planting rates.

ACKNOWLEDGMENTS

This project was supported in part by a USDA NIFA grant, contract number 20186702327682. The authors thank editor Terrance Hurley and the journal's reviewers for their helpful comments.

REFERENCES

Argote, Linda, and Dennis Epple. 1990. "Learning Curves in Manufacturing." *Science* 247(4945): 920–4.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29(3): 155–73.

Assefa, Yared, Paul Carter, Mark Hinds, Gaurav Bhalla, Ryan Schon, Mark Jeschke, Steve Paszkiewicz, Stephen Smith, and Ignacio A. Ciampitti. 2018. "Analysis of Long Term Study Indicates both Agronomic Optimal Plant Density and Increase Maize Yield per Plant Contributed to Yield Gain." *Scientific Reports* 8(1): 1–11.

Assefa, Yared, P.V. Vara Prasad, Paul Carter, Mark Hinds, Gaurav Bhalla, Ryan Schon, Mark Jeschke, Steve Paszkiewicz, and Ignacio A. Ciampitti. 2016. "Yield Responses to Planting Density for US Modern Corn Hybrids: A Synthesis-Analysis." *Crop Science* 56(5): 2802–17.

Benjamin, Daniel J. 2019. "Errors in Probabilistic Reasoning and Judgement Biases." *Handbook of Behavioral Economics* 1(2): 69–186.

Benkard, C. Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90(4): 1034–54.

Bold, Tessa, Kayuki C. Kaizzi, Jakob Svensson, and David Yanagizawa-Drott. 2017. "Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda." *Quarterly Journal of Economics* 132(3): 1055–100.

Chavas, Jean-Paul, Guanming Shi, and Joseph Lauer. 2014. "The Effects of GM Technology on Maize Yield." *Crop Science* 54(4): 1331–5.

Ching, Andrew T., Tülin Erdem, and Michael P. Keane. 2013. "Learning Models: An Assessment of Progress, Challenges, and New Developments." *Marketing Science* 32(6): 913–38.

Ciliberto, Federico, GianCarlo Moschini, and Edward D. Perry. 2019. "Valuing Product Innovation: Genetically Engineered Varieties in U.S. Corn and Soybeans." *RAND Journal of Economics* 50(3): 615–44.

Clancy, Matthew S., and GianCarlo Moschini. 2017. "Intellectual Property Rights and the Ascent of Proprietary Innovation in Agriculture." *Annual Review of Resource Economics* 9: 53–74.

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *American Economic Review* 100(1): 35–69.

Coulter, Jeff. 2021. *Optimal Corn Plant Populations in Minnesota*. St. Paul, MN: University of Minnesota Extension. https://extension.umn.edu/corn-planting/optimal-corn-plant-populations-minnesota

de Bruin, Jason L., and Palle Pedersen. 2008. "Effect of Row Spacing and Seeding Rate on Soybean Yield." *Agronomy Journal* 100(3): 704–10.

DeGroot, Morris H. 2005. *Optimal Statistical Decisions*, Vol 82. Hoboken, NJ: John Wiley & Sons.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2008. "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review* 98(2): 482–8.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101(6): 2350–90.

Duvick, Donald N. 2005. "The Contribution of Breeding to Yield Advances in Maize (*Zea mays* L.)." *Advances in Agronomy* 86: 83–145.

Emerick, Kyle, and Manzoor H. Dar. 2021. "Farmer Field Days and Demonstrator Selection for Increasing Technology Adoption." *Review of Economics and Statistics* 103(4): 680–93.

Erdem, Tülin, and Michael P. Keane. 1996. "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science* 15(1): 1–20.

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

Foster, Andrew D., and Mark R. Rosenzweig. 2010. "Microeconomics of Technology Adoption." *Annual Review of Economics* 2(1): 395–424.

Gaspar, Adam P., Spyridon Mourtzinis, Don Kyle, Eric Galdi, Laura E. Lindsey, William P. Hamman, Emma G. Matcham, et al. 2020. "Defining Optimal Soybean Seeding Rates and Associated Risk across North America." *Agronomy Journal* 112(3): 2103–14.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25: 501–22.

Haigh, Michael S., and John A. List. 2005. "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis." *Journal of Finance* 60(1): 523–34.

Handel, Benjamin, and Joshua Schwartzstein. 2018. "Frictions or Mental Gaps: What's behind the Information We (Don't) Use and when Do We Care?" *Journal of Economic Perspectives* 32(1): 155–78.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning through Noticing: Theory and Evidence from a Field Experiment." *Quarterly Journal of Economics* 129(3): 1311–53.

Henckel, Timo, Gordon D. Menzies, Peter G. Moffatt, and Daniel J. Zizzo. 2021. "Belief Adjustment: A Double Hurdle Model and Experimental Evidence." *Experimental Economics*: 1–42.

Hennessy, David A., Alexander J. Lindsey, Yuyuan Che, Laura E. Lindsey, Maninderpal Singh, Hongli Feng, Elizabeth M. Hawkins, et al. 2021. "Characterizing the Decision Process in Setting Corn and Soybean Seeding Rates." *Journal of Extension* (in press).

Jovanovic, Boyan, and Yaw Nyarko. 1995. "A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves." *Brookings Papers on Economic Activity. Microeconomics* 1995: 247–305.

Lee, Chad D., Dennis B. Egli, and Dennis M. TeKrony. 2008. "Soybean Response to Plant Population at Early and Late Planting Dates in the Mid-South." *Agronomy Journal* 100(4): 971–6.

Lee, Seungki, Yongjie Ji, and GianCarlo Moschini. 2021. *Agricultural Innovation and Adaptation to Climate Change: Insights from Genetically Engineered Maize*." CARD Working Paper 21-WP 616. Ames, Iowa: Iowa State University.

Levitt, Steven D., John A. List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." *Journal of Political Economy* 121(4): 643–81.

List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118(1): 41–71.

Matuschke, Ira, and Matin Qaim. 2009. "The Impact of Social Networks on Hybrid Seed Adoption in India." *Agricultural Economics* 40(5): 493–505.

Moschini, Giancarlo, and David A. Hennessy. 2001. "Uncertainty, Risk Aversion, and Risk Management for Agricultural Producers." *Handbook of Agricultural Economics* 1: 87–153.

Norton, George W., and Jeffrey Alwang. 2020. "Changes in Agricultural Extension and Implications for Farmer Adoption of New Practices." *Applied Economic Perspectives and Policy* 42(1): 8–20.

Olmstead, Alan L., and Paul W. Rhode. 2008. *Creating Abundance*. Cambridge, England: Cambridge Books.

Ortiz-Bobea, Ariel, and Jesse Tack. 2018. "Is another Genetic Revolution Needed to Offset Climate Change Impacts for US Maize Yields?" *Environmental Research Letters* 13(12): 124009.

Perry, Edward D., David A. Hennessy, and GianCarlo Moschini. 2019. "Product Concentration and Usage: Behavioral Effects in the Glyphosate Market." *Journal of Economic Behavior & Organization* 158: 543–59.

Perry, Edward D., and GianCarlo Moschini. 2020. "Neonicotinoids in US Maize: Insecticide Substitution Effects and Environmental Risk." *Journal of Environmental Economics and Management* 102: 102320.

Perry, Edward D., GianCarlo Moschini, and David A. Hennessy. 2016. "Testing for Complementarity: Glyphosate Tolerant Soybeans and Conservation Tillage." *American Journal of Agricultural Economics* 98(3): 765–84.

Phillips, Lawrence D., and Ward Edwards. 1966. "Conservatism in a Simple Probability Inference Task." *Journal of Experimental Psychology* 72(3): 346–54.

Russell, W.A. 1991. "Genetic Improvement of Maize Yields." *Advances in Agronomy* 46: 245–98.

Schwartzstein, Joshua. 2014. "Selective Attention and Learning." *Journal of the European Economic Association* 12(6): 1423–52.

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

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49(2): 326–65.

Thompson, Peter. 2001. "How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study." *Journal of Political Economy* 109(1): 103–37.

Tokatlidis, I.S., and S.D. Koutroubas. 2004. "A Review of Maize Hybrids' Dependence on High Plant Populations and Its Implications for Crop Yield Stability." *Field Crops Research* 88(2–3): 103–14.

Wright, Brian D. 2012. "Grand Missions of Agricultural Innovation." *Research Policy* 41(10): 1716–28.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Perry, Edward D., David A. Hennessy, and GianCarlo Moschini. 2021. "Uncertainty and learning in a technologically dynamic industry: Seed density in U.S. maize." *American Journal of Agricultural Economics* 1–23. https://doi.org/10.1111/ajae.12276