Received: 11 June 2020 | Accepted: 6 December 2021

DOI: 10.1111/ajae.12283

ARTICLE

# On the value of innovation and extension information: SCN-resistant soybean varieties

**Seungki Lee**[1] | **GianCarlo Moschini**[2]

[1]Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, Ohio, USA

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

**Correspondence**
GianCarlo Moschini, Department of Economics, Iowa State University, Ames, IA 50011, USA.
Email: moschini@iastate.edu

**Abstract**

This paper presents direct evidence on the impact of a specific extension program that is aimed at promoting the adoption of varieties resistant to the soybean cyst nematode (SCN), specifically the *Iowa State University SCN-Resistant Soybean Variety Trials*. We use two data sources: experimental data from these variety trials and a rich proprietary dataset on farmers' seed purchases. Combining these data, we estimate the value of soybean cyst nematode-resistant variety availability, and the associated variety trials that provide information on their performance to farmers and seed companies. Given the scope and diffusion of this extension program, the focus of the analysis is on Iowa and northern Illinois over the period 2011–2016. Farmers' seed choices are modeled in a discrete choice framework, specifically a one-level nested logit model. Using the estimated demand model, we find farmers' marginal willingness to pay for soybean cyst nematode-resistant varieties, and for related extension information provided by the Iowa State University SCN-Resistant Soybean Variety Trials program, to be large. These results are confirmed by counterfactual analyses showing that, over the six-year period and region of the study, the total ex post welfare change associated with the existence of, and information about, SCN-resistant seeds is about \$478 million. About one-third of this surplus is captured by seed suppliers, and two-thirds accrues to farmers.

**KEYWORDS**

agricultural extension, nested logit, resistant varieties, seed market, soybean cyst nematode, value of information

**JEL CLASSIFICATION**
O14, O33, Q12, Q16

## 1 | INTRODUCTION

Agricultural extension has long been regarded as an important public service, especially as it relates to technology adoption (Anderson & Feder, 2007; Evanson, 1997). Although the significance of extension is widely recognized, empirical evidence on the magnitude of its economic impact is usually inferred indirectly from estimated links between extension activities and farms' performance or agricultural productivity (e.g., Dinar et al., 2007; Genius et al., 2013; Jin & Huffman, 2016; Maffioli et al., 2011). Studies that provide direct evidence are rarer, and it is recognized that "getting a handle on the value of extension to farmers is not a trivial task" (Anderson & Feder, 2007, p. 2349). In this paper, we provide direct econometric evidence on the impact of a specific extension program, which spans more than two decades and is aimed at promoting the adoption of varieties resistant to the soybean cyst nematode (SCN), specifically the Iowa State University SCN-Resistant Soybean Variety Trials (ISU-SCN). Our analysis is rooted in a structural model of seed demand, which is estimated by leveraging two large and unique data sources.

The soybean cyst nematode is the most harmful pathogen to soybean yields in North America (Allen et al., 2017; Bandara et al., 2020). This plant parasite (a microscopic roundworm) feeds on soybean roots and can result in damages that have severe repercussions on production. Recommended management practices to deal with this pest include crop rotation with non-host plants (such as corn) and, crucially, the adoption of SCN-resistant soybean varieties. Such varieties have been developed over time by including certain wild-type soybeans into the breeding program for commercial varieties. Not all SCN-resistant varieties are equally effective. Resistance is provided by several genes, and it is understood that SCN-resistant varieties can vary a lot in the degree of resistance they possess (and, of course, in their agronomic performance) (Tylka, 2012). Furthermore, there are no regulatory standards that constrain a variety's claim to be SCN-resistant. These considerations have motivated the ISU-SCN program to evaluate hundreds of SCN-resistant varieties each year for the last two decades, providing the most comprehensive set of SCN-resistant soybean variety trials in the nation.

The purpose of this study is to econometrically estimate the value, to farmers and seed companies, of the availability of SCN-resistant varieties and the associated variety trials that provide information on their performance. The presumption of our analysis is that, if the availability of SCN-resistant varieties and knowledge about them and their performance produce value to farmers, then this will be reflected in farmers' choice of seed varieties. In other words, valuable SCN-related characteristics would imply a shift in farmers' seed demand. Whereas simple in principle, the quantitative characterization of this shift is challenging because the analysis needs to be performed at the level on individual seed varieties. This presents two related but distinct challenges: The number of varieties is very large, and information about choices pertaining to individual varieties is needed. A common way to deal with the first issue—modeling demand when the set of products is very large—is to follow the so-called characteristics-based approach in a discrete choice framework, and this is what we do.

To address the second challenge—the availability of suitable data at the individual variety level—our empirical analysis relies on two unique data sources: first, the extensive ISU-SCN variety trials, which span the period 1997 to present. Over this period, ISU-SCN has tested a large number of commercially available SCN-resistant varieties (about 125 soybean varieties per year). Performance metrics from field trials, carried out annually at nine locations in Iowa, include yield rate and end-of-season SCN population density. Results from these trials have been diffused broadly. Importantly, in addition to being freely accessible online, starting from fall 2010 these results have been directly mailed to Iowa and northern Illinois farmers as a supplement to two widely distributed farm magazines. The second data source we utilize is a large proprietary dataset of plot-level seed purchases, by a representative sample of soybean farmers, collected by Kynetec USA, Inc. These data provide variety-level estimates of farmers' choices of soybean seed varieties and are available to us from 1996 to 2016.

The methodology we apply relies on estimating a discrete choice model of farmers' soybean seed demand, similar to the framework of Ciliberto et al. (2019) but with a notable difference. Specifically, given the nature of the research question addressed here, the seed demand model we specify needs to be at a much more granular level, namely at the individual variety level. Because the ISU-SCN program targeted varieties mostly suited to Iowa and northern Illinois, our empirical analysis focuses on seed purchases in Iowa and northern Illinois. Furthermore, as discussed in more details in what follows, information from this program was considerably enhanced starting with the results of 2010 trials, which were made available before the 2011 planting season. Hence, the econometric analysis focuses on seed demand over the period 2011–2016. The analysis is carried out at the market level, where markets are identified by the crop reporting districts (CRD) and the year (Iowa and northern Illinois together are composed of 14 CRDs).

We use a one-level nested logit demand model, where individual soybean varieties are the "inside goods" and the observed acreage of corn grown provides our measure of the "outside option" that defines the potential market size. This structure maintains that, on a given plot, there is higher substitutability between soybean varieties themselves than between soybean and corn varieties (an attractive property in view of the widespread practice of crop rotation). Our estimation procedure, based on Berry (1994), handles standard endogeneity concerns relating to price by the use of instrumental variables. Furthermore, we control for products' life cycles, a potentially confounding factor that is specific to the variety-level specification of demand of this study.

Based on the estimated demand model, we calculate farmers' willingness to pay (WTP) for the SCN resistance trait and the extension information of tested SCN-resistant varieties. Extension information about a given variety is proxied by three metrics: (a) being tested; (b) being tested and performing above the median yield within the test sample; and, (c) being tested and performing above the median SCN-infestation control within the test samples. Results indicate a substantial response to these indicator variables: the WTP for the SCN resistance is \$2.69/acre, and the WTP for an SCN-resistant variety that is tested by ISU-SCN and performs in the top 50% of varieties for the two metrics considered is \$9.62/acre. Estimated WTPs provide a first-order approximation to the total surplus produced by the innovation. The total surplus attributable to the availability of SCN resistant varieties, and the associated information provided by the ISU-SCN program, is estimated at \$466 million (for in Iowa and northern Illinois during 2011–2016).

We provide a fuller assessment of the welfare implications of SCN resistance, and the related extension activities, by using the estimated structural model to assess three counterfactual scenarios: (a) the absence of both SCN resistance traits and the associated ISU-SCN extension program; (b) the absence of the ISU-SCN program only; and, (c) the absence of variety-specific performance metrics produced by the extension program. Prices for the counterfactual scenarios are predicted through a hedonic price regression, as in Hausman and Leonard (2002). In our discrete choice formulation, the expected profit from seed choices can be computed analytically as the inclusive values of the choice set that farmers face. Differences between the inclusive values of alternative scenarios (e.g., with and without ISU-SCN) permit calculation of welfare gains for farmers. The counterfactual demands with predicted prices also allow us to calculate the net revenue change of seed suppliers. Over the period and region of this study, the model predicts about \$478 million of total ex post welfare gains from the joint contribution of SCN resistance availability and the ISU-SCN extension program, with seed suppliers capturing about one-third of this surplus and farmers obtaining two-thirds.

The rest of the paper is organized as follow. We first provide additional background on the SCN and the ISU-SCN program, as well as a description of the ISU-SCN data and the Kynetec seed purchase data. This is followed by a discussion of the modeling framework, including details on product and market definitions. The nested logit, discrete choice model is specified next, and this is followed by a presentation of the estimation results. This is followed by welfare metrics obtained from the estimated model, including counterfactual analyses to tease out the separate value of the availability of SCN-resistant varieties, and the added value of the related extension program.

## 2 | BACKGROUND AND DATA

The value of SCN-resistant varieties to farmers (and society), in our model, arises from two sources: innovation in the seed breeding industry, which has generated commercial varieties resistant to SCN; and, extension activities, specifically the ISU-SCN program, which include experimental test results to verify the effectiveness of individual varieties' resistance, as well as the dissemination of the associated information to farmers.

### 2.1 | SCN and extension information

The SCN (*Heterodera glycines* Ichinohe) has been reported as the most damaging pathogen of soybean in North America for more than two decades (Allen et al., 2017; Bandara et al., 2020; Koenning & Wrather, 2010; Wrather et al., 2001). In the United States, the SCN was first discovered in North Carolina in 1954 and is currently found in all soybean-producing states except West Virginia (Tylka & Marett, 2021). This plant parasite, a microscopic roundworm, feeds on soybean roots and can retard plant growth, causing a serious yield loss (yields are lower because fewer pods develop on infected plants). Because the visual symptoms of SCN damage are hard to observe, farmers may not be fully cognizant of the problem they face, and a major focus of agricultural extension in this setting has indeed been that of improving farmers' awareness by providing objective information.

Two recommended strategies to control this pest are crop rotation with non-host plants (such as corn), and, crucially, the adoption and rotation of SCN-resistant varieties (Niblack, 2005).[1] Such resistant varieties have been developed over time by including certain wild-type resistant lines (the source of resistance) into the breeding program of commercial varieties. This process is lengthy and difficult, due to the complex and polygenic nature of SCN resistance. There is also the concern that resistant varieties may experience a yield penalty when pest pressure is low. As a result, there exists a considerable variation of the SCN resistance between SCN-resistant varieties, and yield performance and SCN population suppression can vary a lot depending on the seed choice as well as land condition and other agronomic factors (Tylka, 2012).

Iowa State University conducts the most comprehensive SCN-resistant variety trials among similar extension programs in the United States (Staton, 2013). In the ISU-SCN, information for extension is procured from field experiments on SCN-resistant varieties, carried out annually at up to nine locations in Iowa. More than 100 SCN-resistant varieties are evaluated every year, along with several popular traditional (SCN-susceptible) varieties that serve as experimental controls in replicated field plots. After the harvest, the experiment records yield rates, and collects the soil samples from each experimental plot to count the SCN population density (eggs/100 cc) at the end of the season, a measure of SCN infestation inversely related to the effectiveness of SCN resistance for the given variety. Consequently, the ISU-SCN's reports display SCN-resistant varieties with their SCN-resistance source and field performances (including yield rate and end-of-season SCN density after harvest).[2] Starting in 1997, their annual summary reports have been posted online to be freely accessible and diffused broadly. In the last two decades, an average of about 125 different varieties

[1]Chemical control through the use of soil nematicides is expensive, not particularly effective, and seldom used. By contrast, the use of nematode-protectant seed treatment is gaining some acceptance in recent years, but such protectants are best viewed as supplementing the primary SCN management strategies (Bissonnette & Tylka, 2017).

[2]Section B of the online supplementary material provides more details about the ISU-SCN data.

each year have been documented. Importantly, since fall 2010, the reports have been directly mailed in the post-harvest season to virtually all Iowa and northern Illinois farmers as a supplement to the magazines *Iowa Farmer Today* and *Illinois Farmer Today*. Through these weekly periodicals the ISU-SCN reports have been distributed, free of charge, to about 70,000 farm owners and operators in Iowa and northern Illinois.[3]

Consistent with the scope of the ISU-SCN program, and the diffusion of extension information, we define the region and time for the study as Iowa and northern Illinois, with seed choices spanning the period 2011 to 2016.

## 2.2 | Seed purchase data

For seed purchase observations, we use a proprietary dataset (TraitTrak) for soybean seed purchases, collected by the survey company Kynetec USA, Inc. These farm-level data provide rich information on plot-level seed purchases such as price, seed trait, variety, brand, parent company, quantity purchased, and projected acres. The Kynetec data are available to us from 1996 to 2016. Based on the information diffusion from ISU-SCN, we mainly exploit the data after 2010. Notwithstanding that, observations prior to 2010 are also used to provide information about the product life cycle of cultivated varieties and for assembling the stock of known SCN-resistant varieties. The Kynetec dataset is designed to be a representative sample of soybean growing farms at the CRD level.[4] Kynetec data for the two states and period of interest (Iowa and northern Illinois over 2011–2016) include an average of 717 farmers per year and 2186 plot-level seed purchase observations per year.

## 2.3 | Descriptive statistics

The ISU-SCN dataset is merged to the Kynetec data at the variety level. Over the period 1997–2015, ISU-SCN tested 1904 varieties (1798 for SCN-resistant varieties, as well as 106 susceptible varieties that served as controls). Not all tested varieties are observed in the seed purchase data: 656 of the SCN-resistant tested varieties are observed in the seed data over the entire period; and, in the estimation period of 2011–2016, 381 tested SCN-resistant varieties are observed.

In this study, we consider two distinct levels of information concerning SCN resistance "attributes" of observed soybean varieties. First, whether a variety is indeed SCN-resistant, that is, it carries genes from the source of resistance genetic stock. We assemble this information from various extension publications,[5] and we treat it as common knowledge as this information is known to seed companies themselves and conceivably conveyed to buyers (farmers) by seed sales agents. The second set of attributes concerns whether a particular SCN-resistant variety was tested by the ISU-SCN program, and the performance metrics resulting from the field trials. Whether or not a soybean variety possesses SCN resistance in its genome may be taken to represent the underlying raw value of innovation, which is brought about by R&D and breeding activities; given that, one would conclude that being tested by the ISU-SCN program and the performance metrics produced and disseminated by this program are valuable information signals that pertain to the true value-added of extension.

[3]Choice of this region is not simply dictated by proximity to the extension service provider, it is also a reflection of relevant agro-climatic conditions. Soybeans are photoperiod sensitive, such that varieties are classified into maturity groups specific to each latitude (Mourtzinis & Conley, 2017). Furthermore, some varieties are better adapted to local growing conditions than others. The varieties included in the ISU-SCN study are most likely to be of interest to farmers in Iowa and northern Illinois.

[4]CRDs are aggregates of counties, as defined by National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA). Section A of the online supplementary material provides a more detailed description on this seed dataset.

[5]In addition to performing field trials on a subset of SCN-resistant varieties, ISU-SCN has endeavored to compile and distribute annually a list of all available SCN-resistant varieties through their so-called PM-1649 publication "Soybean cyst nematode-resistant soybean varieties for Iowa." The total list of all SCN-resistant varieties that we have assembled by combining all extension information, over the period 1997–2016, contains 6912 varieties.

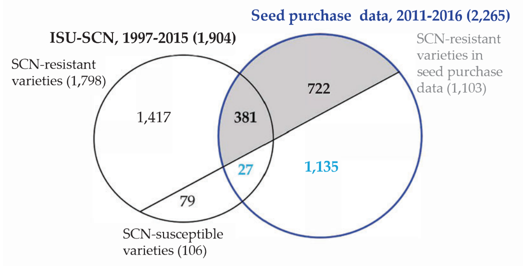


**FIGURE 1** Number of soybean seed varieties: ISU-SCN and Kynetec data, Iowa and northern Illinois. This Venn diagram illustrates the relationship between the number of SCN-resistant and SCN-susceptible varieties, the number of such varieties tested by ISU-SCN, and the number of such varieties observed in the Kynetec seed purchased data

As discussed later, however, more nuanced interpretations of the roots of the value of SCN-related soybean seed characteristics are possible.

Figure 1 provides an overview of our data structure. In the six-year timeframe of this study, in Iowa and northern Illinois, we observe soybean seed purchases for 2265 distinct varieties, among which 1103 are SCN resistant. Within these SCN-resistant varieties, 381 varieties are found in the set of varieties tested by ISU-SCN. Figure 2 provides some evidence on the diffusion and adoption of SCN-resistant varieties over time. From low market shares at the beginning of this period, the uptake of SCN-resistant varieties has been steady, a testament to the commitment of both seed companies, who mustered the required breeding efforts, and extension activities, which educated farmers to recognize and deal with SCN-infested production conditions. Around 40% of soybean acreage is accounted for by ISU-SCN tested resistant varieties in our study period, 2011–2016.

## 3 | MODELING FRAMEWORK

The model we develop is rooted in Berry's (1994) influential formulation, which shows that an individual-level discrete choice problem can be aggregated such that it can be estimated, as a linear model, with market level data and that an estimation procedure can be devised to account for critical endogeneity issues via standard instrumental variables techniques.

As discussed in Richards and Bonnet (2018), discrete choice models are particularly useful for problems that entail a large number of choices and when the focus is on the attributes of goods. Discrete choice modeling of agricultural technology adoption was implemented by Useche et al. (2009) in the context of studying genetically engineered (GE) trait adoption in corn seed demand. They use survey data and, as in Nevo (2001), they include demographic information in their multinomial logit demand model. Useche et al. (2012) apply discrete choice modeling to learning, a context also investigated by Ma and Shi (2015). Because our data do not provide demographic information, we specify the seed demand model at the market level, following Berry (1994). In particular, we specify a nested logit model to improve upon the multinomial logit in terms of producing more realistic substitutability patterns.

### 3.1 | Market definition

We define a market in terms of time–region combination, following Berry et al. (1995). Specifically, the regional level of our analysis is the CRD. The area of our study, Iowa and northern Illinois, contains 14 CRDs (nine in Iowa and five in northern Illinois) and the estimation period encompasses six6 years, from 2011 to 2016, as discussed earlier (thus, our analysis covers 84 markets).

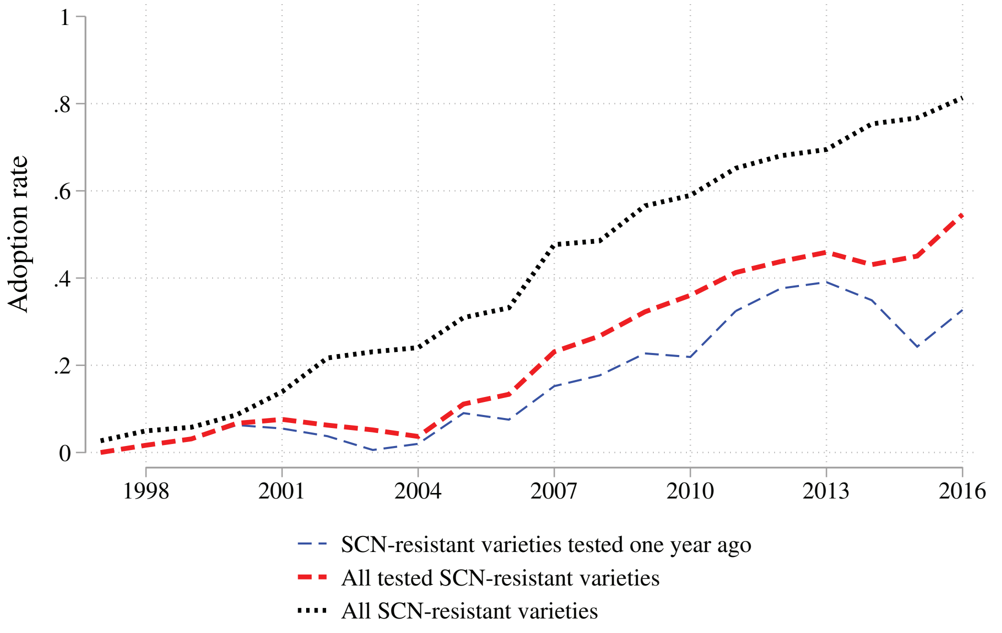


FIGURE 2 Adoption of SCN-resistant soybean seeds, Iowa and northern Illinois, 1997–2016. Each line shows the adoption rate (the portion of soybean acreage planted by the seed group of interest). Three seed groups are explicitly represented: (a) SCN-resistant varieties tested by ISU-SCN 1 year ago; (b) SCN-resistant varieties tested at some point in the past; and, (c) all SCN-resistant varieties *Source*: see text

In discrete choice models of demand, a necessary step concerns the definition of the potential market size. In our context, the relevant market size is the number of acres that are potentially available for soybean planting. We take this to be represented by the total area planted to either corn or soybeans. In the seed purchase sample we use, area planted to either soybeans or corn is, on average, 2.6 million acres per market (41% of which planted to soybeans). These are by far the two most important crops in Iowa and Illinois. Although some corn monoculture is practiced, in most cases farmers follow rotation practices whereby soybeans almost always follow corn, either in a corn–soybean rotation or in a corn-corn-soybean rotation.

## 3.2 | Brands

Table 1 illustrates the supply structure of soybean seeds, for the period and region of interest, at three levels: parent company, brand, and variety (the latter distinguished varieties according to their SCN-related characteristics). A total of 93 distinct seed brands are observed in Iowa and northern Illinois during 2011–2016, but Table 1 lists only individual brands that have more than 1% market share. Five major companies account for about 82% of soybean seed sales, and are listed separately—the other significant brands are shown in the "local and regional" company group. For the purpose of controlling for brand fixed effects, in the regression analysis that follows, brands with less than 1% share are combined in an "others" set within each of their respective company groups.

From Table 1, it is apparent that the soybean seed market configuration is that of a highly concentrated, differentiated-product industry. The two most prominent brands from Dupont and Monsanto (Pioneer and Asgrow, respectively) account for more than half of the market share. From

**TABLE 1** Market shares and number of varieties, Iowa and northern Illinois, 2011–2016

| Parent company—main brands | Market share | Number of varieties | | |
|---|---|---|---|---|
| | | Total | SCN-resistant | Tested by ISU-SCN |
| AgReliant—LG Seeds | 2.66% | 127 | 91 | 11 |
| Dow Agrosciences—Prairie Brand, Mycogen | 4.82% | 230 | 125 | 60 |
| Dupont—Pioneer | 35.82% | 365 | 152 | 81 |
| Monsanto—Asgrow, Channel, Kruger Seeds, Stone Seed Farms | 29.04% | 454 | 257 | 112 |
| Syngenta—NK Seeds | 9.68% | 115 | 66 | 58 |
| Local & regional companies—Beck's Hybrids, Stine Seed Company, Growmark/FS, Croplan Genetics | 17.98% | 974 | 412 | 120 |
| Total | 100% | 2265 | 1103 | 442 |
| of which GT | | 1958 | 964 | 420 |
| of which conventional | | 307 | 139 | 22 |

*Source*: ISU-SCN data and Kynetec data.

**TABLE 2** Summary statistics

| Variable | Mean | *SD* | Min | Max | *N* |
|---|---|---|---|---|---|
| Price ($/acre) | 51.95 | 9.52 | 15.43 | 114.30 | 7141 |
| SCN resistant varieties | 52.65 | 9.01 | 18.74 | 114.30 | 4626 |
| Non-resistant varieties | 50.68 | 10.26 | 15.43 | 105.37 | 2515 |
| GT varieties | 52.70 | 9.16 | 18.74 | 114.30 | 6547 |
| Conventional varieties | 43.68 | 9.43 | 15.43 | 73.41 | 594 |
| SCN resistance trait | 0.65 | 0.48 | 0 | 1 | 7141 |
| Tested by ISU-SCN | 0.36 | 0.48 | 0 | 1 | 7141 |
| Yield top 50% | 0.20 | 0.40 | 0 | 1 | 7141 |
| SCN control top 50% | 0.18 | 0.38 | 0 | 1 | 7141 |
| GT trait | 0.92 | 0.28 | 0 | 1 | 7141 |
| Product age | 3.60 | 3.56 | 1 | 21 | 7141 |
| No. of products per market | 85.01 | 27.82 | 37 | 142 | 84 |
| Inside option market share | 0.41 | 0.06 | 0.26 | 0.56 | 84 |
| Outside option market share | 0.59 | 0.06 | 0.44 | 0.74 | 84 |

*Note*: Prices are expressed in 2011 dollars.

the perspective of demand estimation, equilibrium pricing in such an industry inevitably raises the issue of potentially endogenous prices, which we will address at the estimation stage.

Soybean seed prices are typically quoted in $/bag, where a bag historically contained 50 lbs of seed. Starting in 2013, the industry moved to units defined by seed count, with a bag containing 140,000 seeds. For clarity, in this article seed prices are expressed as seed expenditure per planted acre. Table 2 reports some summary statistics of the data used in estimation, including prices.[6] Prices are reported separately for SCN-resistant and susceptible varieties, as well as for conventional and glyphosate tolerant (GT) groups. Although GT products are generally about $9/acre more expensive

[6]Consistent with the empirical application that follows, nominal prices are deflated by the crop sector index for price paid, as provided by USDA (index = 1 in 2011).

than non-GT products, SCN resistance does not appear to command large price premiums. The average difference between SCN-resistant and regular (susceptible) varieties is about $2/acre.

## 3.3 | Products and seed traits

In the discrete choice model that we employ for the empirical analysis, the choice set must satisfy three characteristics: mutually exclusive, exhaustive, and a finite number of alternatives (Train, 2009). Farmers can choose only one seed product in a given plot, so the choice situation obviously fits the discrete choice framework, and as long as the three conditions above are satisfied, any way of product definition can technically work. The seed demand model of Ciliberto et al. (2019) relies on the notion of "product lines," defined by a combination of four components: crop (corn or soybeans), parent company, brand, and presence of GE traits. Our product definition needs to be more refined than that, however, because the nature of our research question requires products to be defined at the variety level. In total, over all markets considered, we have 2265 distinct varieties.

A major trend in the soybean industry over the last two decades has been the adoption of GT varieties, that is, GE varieties that can withstand over-the-top application of the broad-spectrum herbicide glyphosate. GT soybean varieties were rapidly adopted, following their introduction in 1996, and this diffusion process reached its maturity around 2010, when the share of U.S. soybean acres planted with GT seeds plateaued at around 93% (see, e.g., Fernandez-Cornejo et al., 2014). This observation provides an additional justification for our choice to focus the seed demand model over the 2011–2016 period. To be specific, during the GE adoption phase, it is likely that a choice of SCN-resistant variety could have happened incidentally, not because of ISU-SCN but because of the GE trait. Post 2010, however, after the adoption of soybean GT varieties had reached a plateau, this issue should not affect our analysis.

Table 2 shows the average number of varieties in each market over the period of study. It is apparent that the choice sets in our model are quite large, including an average of about 85 varieties per market. This table also reports the contemporaneous standard deviation for the number of varieties (across CRDs), which illustrates a fair amount of choice-set variation. The data also show a resurgence of conventional varieties in more recent years, from an average 3.8 per market in 2011 to 14 per market in 2016, likely in response to the emergence of glyphosate-resistant weeds.

## 3.4 | Product life cycle

Because a distinguishing feature of our analysis is that it is carried out at the variety level, we also need to account for the "product life cycle," an issue that did not arise in the "product line" definition of products used by Ciliberto et al. (2019). Specifically, seed companies continuously introduce new varieties and discontinue old varieties, and newly released seed varieties tend to have a relatively short life cycle (Magnier et al., 2010). Because in our demand model the desirability of a variety is reflected in its market share, and the latter in turn is influenced by its life cycle, ignoring the product life cycle would heavily bias estimation results. We characterize this attribute of a variety by its "age," defined as the number of years since its first market introduction. To determine the latter we use the entire sample in the Kynetec data, which encompasses 31 states (not just Iowa and Illinois) and 21 years.[7] For varieties observed chosen by Iowa and northern Illinois farmers in our sample, the average age is 3.60 years, and the average life cycle (i.e., the number of years a variety is observed in the full U.S. sample) is 6.88 years.

[7]Inevitably, some truncation arises earlier in the sample—varieties grown in 1996, the first year in our sample, are assumed to be introduced in that year. This earlier truncation effect tends to wash out over time, and it is insignificant for the period 2011–2016 used in the econometric analysis.

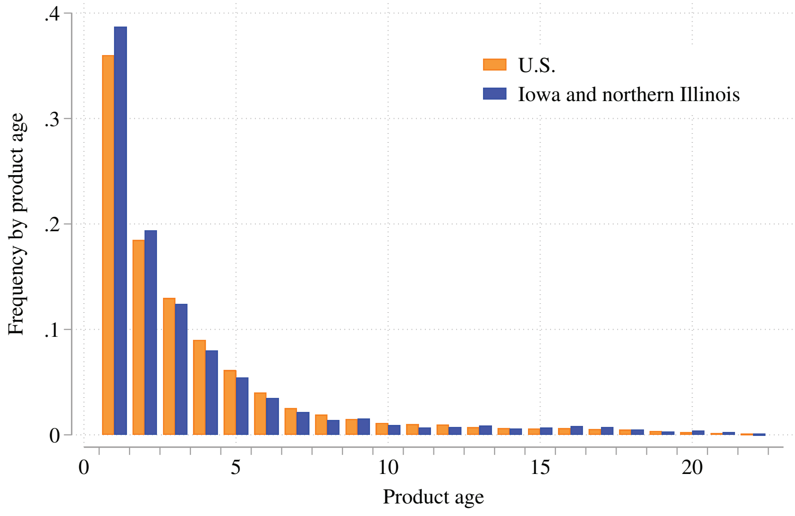


**FIGURE 3** Obsolescence pattern of soybean seeds, 2011–2016. This chart illustrates the frequency of soybean seed varieties by age. Frequency is computed using the entire U.S. soybean seed sample (orange) and the subsample of Iowa and northern Illinois (blue), with age computes with information starting in 1996

Figure 3 illustrates the frequency of soybean seed products by their commercial age, both for the region of interest (Iowa and northern Illinois) and the entire United States. We report frequencies instead of market share by product age because frequencies may not be as affected by market-related circumstances such as price or seed trait. It is apparent that there exists a strong obsolescence pattern in soybean seed varieties. For example, during the period of 2011 to 2016, varieties commercialized within three years accounted for about 70% of the available seed products. Conditioning the demand model by a variety's age, therefore, accounts for the average handicap faced by older varieties in the market and thus aids in the identification of the other determinants of seed choice that are of interest.

## 4 | SEED DEMAND MODEL

Consider a farmer choosing the seed variety $j$, on plot $i$, in market $m$, and denote the expected per-acre profit associated with each element of the farmer's choice set as $\pi_{ijm}$. The farmer is assumed to choose the seed variety that provides the largest expected profit. Given that $J_m$ varieties are available in market $m$, the farmer's profit-maximizing choice entails solving

$$\max_j \pi_{ijm}, \quad j \in \{0, 1, \ldots, J_m\} \tag{1}$$

where $j = 0$ is the outside option (i.e., growing corn).

To implement this discrete choice problem, one needs to parameterize $m$ the payoffs $\pi_{ijm}$. Following standard practice, we write the per-acre payoffs as linear functions of observable attributes and unobservable components, as follows:

$$\pi_{ijm} = \beta \cdot p_{jm} + \gamma \cdot x_j + \sum\nolimits_{k=1}^{9} \alpha_k Z^k_{jt[m]} + \xi^A_{jt[m]} + \xi_{t[m]} + \xi_{l[m]} + \xi_{b[j]} + \xi_{jm} + \varepsilon_{ijm} \tag{2}$$

where $p_{jm}$ denotes the (deflated) variety's price, expressed on a per-acre basis, and $x_j$ is a dummy variable that codes for whether or not the variety in question includes the GT trait.

The next set of variables in (2) reflects information related to SCN resistance and the ISU-SCN program. We capture such information in terms of four primitive indicator variables. The first such variable, $I_j^{SCN-r}$, takes the value of 1 if the given variety possesses SCN-resistance, and 0 otherwise. Next, the indicator variable $I_{j,t[m]}^{tested}$ takes the value of 1 if the given variety had been tested by ISU-SCN (and the results disseminated) by the year defined by market $m$, and 0 otherwise. Finally, other qualitative results of ISU-SCN experiments are capture by two performance indicators: $I_{j,t[m]}^{top-y}$ flags varieties that performed better than the 50th percentile in terms of yield, and $I_{j,t[m]}^{top-s}$ flags varieties that performed better than the 50th percentile in terms of end-of-season SCN population density (a measure of SCN infestation and thus a metric of SCN resistance).[8] Descriptive statistics on the distribution of these four indicator variables are reported in Table 2.

Given these four "treatments," varieties in any given market fall into 10 mutually exclusive groups. Taking "not-SCN-resistant, not tested" as the reference group, the treatments are fully identified by the nine suitably defined dummy variables that appear in Equation (2)[9]:

$$Z_{jt[m]}^1 \equiv I_j^{SCN-r}$$

$$Z_{jt[m]}^2 \equiv I_j^{SCN-r} \times I_{j,t[m]}^{tested} \times I_{j,t[m]}^{top-y} \times I_{j,t[m]}^{top-s}$$

$$Z_{jt[m]}^3 \equiv I_j^{SCN-r} \times I_{j,t[m]}^{tested} \times I_{j,t[m]}^{top-y}$$

$$Z_{jt[m]}^4 \equiv I_j^{SCN-r} \times I_{j,t[m]}^{tested} \times I_{j,t[m]}^{top-s}$$

$$Z_{jt[m]}^5 \equiv I_j^{SCN-r} \times I_{j,t[m]}^{tested}$$

$$Z_{jt[m]}^6 \equiv I_{j,t[m]}^{tested} \times I_{j,t[m]}^{top-y} \times I_{j,t[m]}^{top-s}$$

$$Z_{jt[m]}^7 \equiv I_{j,t[m]}^{tested} \times I_{j,t[m]}^{top-y}$$

$$Z_{jt[m]}^8 \equiv I_{j,t[m]}^{tested} \times I_{j,t[m]}^{top-s}$$

$$Z_{jt[m]}^9 \equiv I_{j,t[m]}^{tested}.$$

The remaining terms in Equation (2) include a set of fixed effects meant to control for variables that affect the per-acre profit but are unobserved (such as other input prices). Specifically, $\xi_{jt[m]}^A$, $\xi_{t[m]}$, $\xi_{l[m]}$, and $\xi_{b[j]}$ represent the product age, time (year), region (CRD), and brand fixed effects; the term $\xi_{jm}$ captures all other unobserved product-market specific components; and, the term $\varepsilon_{ijm}$ represents elements that are specific to plot $i$ and variety $j$ that are unobservable to the researcher but are known to the farmer making the seed choice. The parameters to be estimated are $\beta$, $\gamma$, and the set of $\alpha_k$ (and, of course, all of the included fixed effects).

[8]These performance metrics measure a variety's specific performance (for yield or end-of-season SCN density) relative to the experiment's control group (non-resistant varieties). Section B of the online supplementary appendix provides more details. Results for an alternative specification of performance metrics are reported in Section F of the online supplementary appendix.
[9]This is illustrated in Table X4 in the online supplementary appendix.

## 4.1 | Nested logit

To make the foregoing discrete choice framework operational, we need to make assumptions about the distribution of unobserved terms. Following Berry (1994), it is convenient to represent with $\delta_{jm}$ all terms in Equation (2) that are common to all plots in the same market and planted with the same variety. That is, $\delta_{jm}$ represents the mean expected per-acre profit of variety $j$ in market $m$. Hence, per-acre profit can be represented as:

$$\pi_{ijm} = \delta_{jm} + \varepsilon_{ijm} \tag{3}$$

From the problem in Equations (1) and (2), and the structure in Equation (3), observing the selection of variety $j$, in a given choice situation, means that $\varepsilon_{ijm} - \varepsilon_{ikm} \geq \delta_{km} - \delta_{jm}$. Assuming that the unobserved terms $\varepsilon_{ijm}$ are identically and independently drawn from a Type I Extreme Value (TIEV) distribution, then choice probabilities (or, equivalently, market shares) take the familiar multinomial logit structure (Train, 2009):

$$s_{jm} \equiv \Pr\left(\pi_{ijm} \geq \pi_{ikm}, \forall k \neq j\right) = \frac{\exp\left(\delta_{jm}\right)}{\sum_k \exp(\delta_{km})}. \tag{4}$$

It is well known, however, that the multinomial logit model entails unrealistic substitution patterns (e.g., Debreu, 1960). For example, suppose that one soybean variety becomes unavailable. The multinomial logit model would imply the farmer would be equally likely to choose the outside option (corn) as any other soybean product to replace the discontinued product, whereas one would expect other soybean varieties to be closer substitutes. To deal with this issue, we apply a one-level nesting structure. As discussed earlier, for a given plot farmers first choose either the outside option (corn) or the inside option (soybeans).[10] Conditional on planting soybeans, farmers select a specific seed variety. We also note that this nesting structure is consistent, inter alia, with the common practice of crop rotation.

This one-level nesting model puts structure on the plot-specific unobserved component, $\varepsilon_{ijm}.\forall k \in \{0,1,\ldots,J_m\}.$ Having grouped individual choices as inside option and outside option, we denote these two exclusive groups by $g \in \{0,1\}$, where $g = 0$ means a farmer plants corn, whereas $g = 1$ indicates choosing one of the soybean varieties. Given that, the unobserved component is written as:

$$\varepsilon_{ijm} = v_{igm} + (1-\sigma)v_{ijm} \tag{5}$$

where $v_{ijm}$ is independent and identically drawn from a TIEV distribution; $v_{igm}$ is a term that is common to all varieties in the group; and, the nesting parameter $\sigma \in [0,1)$ captures correlation between varieties within the inside option group. The term $v_{igm}$ is assumed to have the unique distribution such that $\varepsilon_{ijm}$ again follows the TIEV distribution (Cardell, 1997). The larger the $\sigma$, the stronger the correlation between the varieties within the group. That is, if $\sigma$ is significantly high, the nesting structure becomes compelling and farmers tend to stay in the inside option when switching to another choice. By contrast, as $\sigma$ approaches 0 the model reduces to the simple multinomial logit.

One of the attractive properties of the nested logit model is that, although generalizing the substitution pattern between alternatives, relative to the multinomial logit, still yields a closed-form representation of choice probabilities with the nesting structure (Berry, 1994; Train, 2009). In market $m$, the conditional share of variety $j$ (one of the inside options) is:

[10]Saved soybean seeds are also included in the outside option as they are not commercially traded. In our data, about 1.68% of soybean farmland per market are observed using saved seeds.

$$s_{jm|g=1} = \frac{\exp\left(\frac{\delta_{jm}}{1-\sigma}\right)}{\sum_{k \in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)} \tag{6}$$

where $J_m^1$ is the set of soybean products only (e.g., for $g = 1$) in market $m$. Without loss of generality, for the outside option we set $\delta_{0m} = 0$, implying $\pi_{i0m} = \varepsilon_{i0m}$. The probabilities of choosing the inside option ($S_{1m}$), and outside option ($s_{0m}$), are, respectively:

$$S_{1m} = \frac{\left[\sum_{k \in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{(1-\sigma)}}{1 + \left[\sum_{k \in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{(1-\sigma)}}, \tag{7}$$

and

$$s_{0m} = \frac{1}{1 + \left[\sum_{k \in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{(1-\sigma)}}. \tag{8}$$

The unconditional probability of choosing a soybean variety $j$ can then be defined as $s_{jm} \equiv s_{jm|g=1} \cdot S_{1m}$. Using (6), (7), and (8), it then follows that $\ln\left(s_{jm}/s_{0m}\right) \equiv \delta_j + \sigma \ln\left(s_{jm|g=1}\right)$. Recalling the structure of the mean profit terms $\delta_j$—implicitly defined by the identity of (2) and (3)—the log ratio of market shares is:

$$\ln\left(s_{jm}/s_{0m}\right) = \beta p_{jm} + \gamma x_j + \sum_{k=1}^{9} \alpha_k Z_{jt[m]}^k + \sigma \ln\left(s_{jm|g=1}\right) + \xi_{jt[m]}^A + \xi_{t[m]} + \xi_{l[m]} + \xi_{b[j]} + \xi_{jm} \tag{9}$$

Thus, the parameters of the structural demand model can be recovered by estimating this linear regression.

## 4.2 | Identification

The key identification issue, in this setting, is related to the possible price endogeneity in Equation (9) and also the endogeneity of the conditional share appearing on the right-hand side of Equation (9). As shown in Table 1, the market is highly concentrated, products are differentiated by known attributes, and presumably the observed seed variety prices display the equilibrium price choices of seed firms. In such a setting, one should expect a positive correlation between $\xi_{jm}$ and $p_{jm}$ in Equation (9). This is because the term capturing product and market-specific attributes of seed variety $j$, which is unobserved to the econometrician, is arguably known to firms when they make their pricing decisions. Without controlling for this correlation, the price coefficient $\alpha$ would be biased.

To deal with the foregoing identification issues, we follow standard practice and assume that the location of products in the product space (i.e., the distinguishing characteristics of commercial varieties, in our case) is exogenous to the pricing decisions, a strategy originally suggested by Bresnahan (1987) and developed and implemented by Berry (1994) and Berry et al. (1995). We follow Ciliberto et al. (2019) and use functions of the traits in competing varieties as instrumental variables (IVs). They provide a detailed discussion of why the assumption of exogenous location in product space may be particularly reasonable for the seed industry. This is because the introduction of new varieties, especially those embedding special traits, takes a long time, is affected by stochastic elements, and is arguably largely exogenous to firms' pricing decisions. In particular, we use two such IVs: the number of competing products in the same market that share the same trait (i.e., either conventional or GT), both regardless of brand and outside the variety's own brand. In addition, we also use the previous year's corn acreage in that CRD as an instrument. This is meant to capture both a market size effect and the potential impact of crop rotation. Rotation between corn and soybeans plays an important role in midwest agriculture, although its dynamics can be complex, especially in the short run (Hendricks et al., 2014; Kim & Moschini, 2018).

## 5 | ESTIMATION RESULTS

Table 3 reports estimation results for the demand model. Results from both ordinary least squares (OLS) and two-stage least squared (2SLS) that rely on instrumental variables are reported. Our main interest, of course, is in the 2SLS estimates; OLS results are presented for comparison purposes to help assess the instrumental variable procedure we implement. We note at the outset that estimates for $\alpha_6$ and $\alpha_8$ are not reported because these parameters cannot be identified by the data: The corresponding treatment groups (SCN-susceptible and tested varieties that performed in the top group for both performance metrics, and for the SCN control metric) turned out to be empty (Table X4 in the online supplementary appendix).

OLS estimates are clearly problematic. Most worryingly, the nesting parameter $\sigma$ appears close to violating its theoretical upper limit.[11] Not surprisingly, then, virtually all parameters (including the price coefficient) turn out to be not significantly different from zero. The 2SLS estimates perform much better. The use of instrumental variables, as discussed earlier, is meant to account for the endogeneity of prices and of the conditional market shares, $\ln(s_{jm|g=1})$. In particular, we note that the estimated price coefficient has the expected negative sign and is larger in magnitude than with OLS, which suggests a positive role of the IVs employed (endogeneity typically biases the price coefficient toward zero).

The other estimated 2SLS coefficients in Table 3 also appear more reasonable. As a benchmark, the GT trait coefficient is large and significantly different from zero, indicating that farmers' per-acre profit is, on average, positively affected by this GE trait (consistent with empirical evidence from much previous work). The nesting coefficient $\sigma$ is now estimated at about 0.81. Although somewhat larger than in many applications, nesting coefficients of this magnitude are not unique (e.g., An & Zhao, 2019) and are well within the theoretical bounds $[0,1)$. This relatively large coefficient can be rationalized by the specifics of our application. Recall that this parameter controls the within-group substitutability of soybean varieties relative to the outside option. Insofar as a soybean variety is a much closer substitute to another soybean variety than it is to corn (because of the farmer's desire to meet rotation objectives, say), then one would expect a large estimate for $\sigma$.[12] Several of the coefficients of the SCN-related indicator variables, specifically of interest in this study, are also significantly different from zero.

To assess the quality of the IVs used, the first-stage regression results are reported in Table 4. All three IVs are statistically significant in the price equation, and one of them is significant in the conditional share equation. As noted by Pakes (2003), given the Bertrand-Nash competition assumption, these coefficients capture equilibrium relationships and may be difficult to interpret, per se. Still, from these estimated coefficients we may note that, holding the total number of competing varieties constant, an increase in the number of products outside of the own brand (more competition) would tend to reduce price, whereas an increase in the number of products of the same brand (less competition) would tend to increase price. A larger market size (as captured by the lagged corn acres) also seems associated with lower price and smaller individual conditional shares.

[11] As noted by a reviewer, $\sigma = 1$ on the left-hand-side of the regression can be re-expressed as $\ln(s_{jm}) - \ln(s_{0m}) - \ln(s_{jm|g=1}) = \ln(S_{1m}) - \ln(s_{0m})$, the model would attempt to identify parameters from the ratio of market shares of soybeans to corn, which does not vary across varieties.

[12] See Kovo and Eizenberg (2017) for further discussion on the issue of identification when the size of the nesting coefficient is close to one.

**TABLE 3** Estimated parameters of the demand model

| $\ln(s_{jm}/s_{0m})$ | | OLS Coefficient | OLS S.E. | 2SLS (IV) Coefficient | 2SLS (IV) S.E. |
|---|---|---|---|---|---|
| $\beta$ | Price ($/acre) | 0.0000007 | (0.000099) | −0.00953* | (0.00504) |
| $\alpha_1$ | SCN resistant (SCN) | 0.00278 | (0.00239) | 0.0256*** | (0.00806) |
| $\alpha_2$ | SCN × Tested × top-y × top-s | −0.00225 | (0.00669) | −0.000400 | (0.0207) |
| $\alpha_3$ | SCN × Tested × top-y | −0.0306 | (0.0363) | 0.139*** | (0.0523) |
| $\alpha_4$ | SCN × Tested × top-s | 0.00488 | (0.00515) | −0.00222 | (0.0160) |
| $\alpha_5$ | SCN × Tested | −0.00789 | (0.00724) | −0.0863*** | (0.0255) |
| $\alpha_7$ | Tested × top-y | 0.0326 | (0.0360) | −0.102** | (0.0452) |
| $\alpha_9$ | Tested | 0.00495 | (0.00582) | 0.119*** | (0.0263) |
| $\gamma$ | GT trait | 0.00950*** | (0.00318) | 0.0924** | (0.0436) |
| $\sigma$ | Nesting corr. | 0.999*** | (0.000786) | 0.810*** | (0.0336) |
| | Constant | −0.897*** | (0.00770) | −1.548*** | (0.262) |
| | Product age FE | Y | | Y | |
| | Year FE | Y | | Y | |
| | CRD FE | Y | | Y | |
| | Brand FE | Y | | Y | |
| | IVs (3) | | | Y | |
| | Observations | 7141 | | 7141 | |
| | R2 | 0.996 | | 0.964 | |

*Note*: Parameters $\alpha_6$ and $\alpha_8$ are not identified by the data at hand (the corresponding sets of treatments are empty). Standard errors are in parentheses.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

The full set of results pertaining to individual varieties' age fixed effects is reported in Section C of the online supplementary appendix. These results illustrate the fact that, ceteris paribus, newer varieties (particularly ages 2 to 4) are positively related to market share. Also, these varieties' age fixed effects appear quite relevant as instrumental variables in first-stage regression (in particular, age is inversely related to varieties' price).

Concerns about the quality of these instrumental variables, and in particular the question of potentially weak instruments, are inevitable in this setting. To investigate this issue, given that our model entails two endogenous variables, we use the conditional F-statistic suggested by Sanderson and Windmeijer (2016) (SW F-statistic in Table 4). Both computed SW F-statistics, for price and conditional market share, appear above 10, the traditional Stock and Yogo (2005) threshold. This is somewhat reassuring, although, again, it is difficult to completely resolve such IV issues.

## 5.1 | Hypothesis tests and willingness-to-pay estimates

The hypothesis test outcomes in Table 5 provide more specific insights from the estimated results. First, the hypothesis that the SCN resistance attribute has no effect on seed demand, requiring $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$, is clearly rejected ($p$-vale of 0.0051). The overall impact of being tested by ISU-SCN is captured by the null hypothesis $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_7 = \alpha_9 = 0$, which is again rejected ($p$-value of 0.0001). Performance indicators provide differing results, however. Scoring in the top 50% in terms of yield has a significant effect—the null hypothesis of no impact, requiring $\alpha_2 = \alpha_3 = \alpha_7 = 0$, is rejected ($p$-value of 0062). Scoring in the top 50% in terms of end-of-season

TABLE 4 First stage regression result

| | Price | | ln(conditional share) | |
|---|---|---|---|---|
| | Coefficient | S.E. | Coefficient | S.E. |
| IVs: No. of competing products | | | | |
| Same trait regardless of brand | 0.135*** | (0.0317) | −0.0109*** | (0.00393) |
| Same trait outside of own brand | −0.147*** | (0.0322) | 0.00319 | (0.00400) |
| IV: Lagged corn acreage | −0.00403** | (0.00163) | −0.000142 | (0.000200) |
| SCN resistant (SCN) | 0.528* | (0.299) | 0.0917*** | (0.0319) |
| SCN × Tested × top-y × top-s | 1.104 | (0.692) | −0.0452 | (0.0940) |
| SCN × Tested × top-y | −4.221 | (5.484) | 1.132** | (0.556) |
| SCN × Tested × top-s | −1.160** | (0.535) | 0.0179 | (0.0713) |
| SCN × Tested | 0.257 | (0.736) | −0.414*** | (0.101) |
| Tested × top-y | 4.035 | (5.461) | −0.942* | (0.552) |
| Tested | 0.643 | (0.600) | 0.562*** | (0.0843) |
| GT trait | 8.980*** | (0.930) | 0.586*** | (0.105) |
| Constant | 48.62*** | (3.465) | −5.088*** | (0.424) |
| $N$ | 7141 | | 7141 | |
| R2 | 0.179 | | 0.190 | |
| F-statistic (3, 7085) | 9.52 | | 13.60 | |
| SW F-statistic (2, 7085) | 14.25 | | 20.29 | |

*Note*: Standard errors are in parentheses. The lagged corn acreage is measured in thousands of acres. Both equations include product age fixed effects, year fixed effects, CRD fixed effects, and brand fixed effects.
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE 5 Hypothesis tests

| Testing | Result |
|---|---|
| Test 1: SCN vs. no SCN | $\widehat{\chi} = 16.69$, Prob $> \chi(5) = 0.0051$ |
| Test 2: Testing vs. no testing | $\widehat{\chi} = 26.95$, Prob $> \chi(6) = 0.0001$ |
| Test 3: Top 50, yield | $\widehat{\chi} = 12.39$, Prob $> \chi(3) = 0.0062$ |
| Test 4: Top 50, SCN control | $\widehat{\chi} = 0.06$, Prob $> \chi(2) = 0.9697$ |

SCN density measure, however, appears to have no impact at all on seed demand—the corresponding null hypothesis $\alpha_2 = \alpha_4 = 0$ fails to be rejected ($p$-value of 0.97).

The magnitude of the estimated parameters concerning product characteristics can be better understood by expressing them in terms of (marginal) WTP. As a benchmark, we can start with the model's estimates of the WTP for the GT trait, which can be obtained by dividing the attribute's parameter by the price coefficient (i.e., $-\gamma/\beta$) (Train, 2009). We find that this WTP is about \$10/acre. This estimate is comparable, albeit somewhat lower, to that reported by Ciliberto et al. (2019), but account must be taken that the region of analysis and the period studied are different.[13] Concerning attributes related to SCN and ISU-SCN information, Table 6 reports three WTP estimates. For the SCN resistance trait per se, estimated as $-\alpha_1/\beta$, we find that WTP equals \$2.69/acre. For

[13]In particular, they study seed demand over 1996–2011, whereas we focus on the more recent 2011–2016 period. Factors possibly contributing to declining farmers' WTP for GT over the period of this study may include declining commodity prices and the emergence of glyphosate resistant weeds.

**TABLE 6** Marginal willingness-to-pay estimates ($/acre)

| Seed attribute | Value | As percent of average seed cost |
|---|---|---|
| SCN resistance trait | 2.69 | 5.2% |
| SCN resistance & tested by ISU-SCN | 6.08 | 11.7% |
| SCN resistance & tested by ISU-SCN & top yield & top SCN control | 9.62 | 18.5% |
| GT trait | 9.69 | 18.7% |

*Note*: Percent figures in the last column are computed relative to the mean price of all soybean varieties in the sample ($51.95/acre).

varieties that are SCN resistant and also included in the ISU-SCN trials, estimated as $-(\alpha_1+\alpha_2+\alpha_3)/\beta$, we find that WTP equals $6.08/acre. Finally, for SCN-resistant varieties that are included in the ISU-SCN field trials and that score in the top 50% of the two performance indicators used in this study, estimated as $-\sum_i \alpha_i/\beta$, the WTP is $9.62/acre.

It is apparent that these estimated WTPs are rather large. For example, the WTP for top performing SCN-resistant varieties is equivalent to about 18% of the average cost of soybean seed in the sample, and it is comparable in magnitude with the WTP for the GT trait. SCN resistance per se, apart from the information associated with ISU-SCN trials, appears to be valued at about 5% of the average cost of seed.

## 5.2 | Elasticities

It is of some interest to express our estimates in terms of elasticities. The own-price and cross-price elasticity for our one-level nested logit model can be computed as follows (for notational simplicity we drop the market subscript; see Björnerstedt and Verboven (2016) for more general formulas):

$$e_{jj} \equiv \frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} = \beta p_j \left[\frac{1}{1-\sigma} - \left(\frac{\sigma}{1-\sigma}\right)\frac{s_j}{s_0} - s_j\right] \quad , \quad j \in \{1,2,\ldots,J\} \tag{10}$$

$$e_{jk} \equiv \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = -\beta p_k \left[\left(\frac{\sigma}{1-\sigma}\right)\frac{s_k}{s_0} - s_k\right] \quad \text{for } j \neq k \in \{1,2,\ldots,J\} \tag{11}$$

$$e_{0j} \equiv \frac{\partial s_0}{\partial p_j}\frac{p_j}{s_0} = -\beta p_j s_j \quad \text{for } j \in \{1,2,\ldots,J\}. \tag{12}$$

We find that the average own-price elasticity is $e_{jj} = -2.59$, which indicates that soybean seed demand at the variety level is quite elastic. This is not surprising, considering the large number of substitutes that are present in farmers' choice sets. For cross-price elasticity within the nest we find, on average, $e_{jk} = 0.015$, and for the cross-price elasticity across nests (between the outside option and a soybean variety) the average is $e_{0j} = 0.00005$. Consistent with the motivation for the nested model specification, a farmer moving away from a given soybean variety is much more likely to purchase another soybean variety, rather than using the outside option (corn).

Beyond these variety-level seed demand elasticities, of interest to firms for their pricing decisions, for other policy-related questions one may be interested in the "aggregate" elasticity of seed demand. The estimated discrete choice model can provide evidence for that as well. Recalling that the market share of all soybean seeds is denoted by $S_1$, the question here is the effect of scaling the prices of all soybean seed varieties, holding the value of the outside option (the price of corn seeds) constant. Thus, the aggregate elasticity of interest can be stated as:

$$e \equiv \frac{\partial S_1(k\mathbf{p})}{\partial k}\frac{k}{S_1}\bigg|_{k=1}$$

where $\mathbf{p}$ is the vector of all soybean seed prices. Because by definition $S_1 \equiv 1 - s_0$, then $\partial S_1/\partial k = -\partial s_0/\partial k$, and the aggregate elasticity satisfies

$$e = -\frac{s_0}{(1-s_0)}\sum_{j=1}^{J} e_{0j}. \tag{13}$$

By utilizing (12), this elasticity can be conveniently rewritten as $e = \beta s_0 \overline{p}$, where $\overline{p} \equiv \sum_j (s_j/S_1)p_j$ is the weighted average of all soybean seed prices. We find that the average value of this elasticity, computed over all markets in our study, is $e = -0.29$. When viewed from the aggregate input demand level, therefore, soybean seed demand is rather inelastic. The fact this elasticity is greater than zero in absolute value, though, suggests that the extent of substitutability between soybean and corn, as captured by the outside option in our model, is not negligible.[14]

## 5.3 | Randomization tests

How much trust can we put on the econometric results concerning the effects of the "treatments" of interest, SCN resistance and evidence from the ISU-SCN trials? Here we present some "placebo" tests that provide support for our results. These tests are in the spirit of randomization inference that has long been helpful to understand causality in statistics (Ho & Imai, 2006). Specifically, we follow Rotemberg (2019) and devise permutation tests to investigate the plausibility of our regression results if, in fact, there were no true impacts of the ISU-SCN program (or SCN resistance trait). The underlying idea of the randomization exercises we implement is to permute the treatment of interest so that it is randomly assigned, maintaining the same proportion of the treated group as the original sample (Berry & Fowler, 2021; Rotemberg, 2019).

We carry out two randomization experiments. In the first, varieties' attributes concerning the ISU-SCN trials—that is, the indicator variables associated with "tested by ISU," "yield in top 50%," and "SCN control in top 50%"—are randomly assigned. This is done while preserving the relative proportions of the treatments of interest. For example, this means that the randomly assigned treatment of being tested by ISU-SCN is drawn from within the group of SCN resistant varieties. In this first randomization experiment, the indicator variable for the attribute "SCN resistance" is kept true to the data. In the second randomization experiment, all SCN-related indicator variables are randomly assigned, again making sure that the assigned treatments are consistent with the original data structure.

After each permutation draw, the 2SLS regression is run. In so doing, we can assess the effect of SCN-related pseudo treatments under the same number of varieties featured by ISU-SCN or SCN resistance trait. Using Monte Carlo simulations, we repeat this procedure 1000 times and report the mean and standard deviation of estimated coefficients. The results of these placebo tests are reported in Table 7. In Randomization 1, the permutation process only affects ISU-SCN terms, whereas the SCN attribute is true to the data. The results show that the attributes subject to randomization (also labeled by "R" in Table 7) are now statistically indistinguishable from zero, whereas the attribute "SCN resistance," which is true to the data in this experiment, remains positive and statistically significant. In Randomization 2, both SCN resistance and ISU-SCN terms are permuted. Here we find

[14]The inclusion of the outside option (corn) also permits us to measure the impact of SCN resistance and ISU-SCN on the extensive margin—see section D in the online supplementary appendix.

**TABLE 7** Distribution of demand estimates under randomization experiments

| | Original 2SLS Estimate | | Randomization 1 | | | Randomization 2 | |
|---|---|---|---|---|---|---|---|
| | | | Mean | *SD* | | Mean | *SD* |
| Price | −0.00953* | | −0.0097 | 0.0001 | | −0.0101 | 0.0002 |
| SCN resistant (SCN) | 0.0256*** | | 0.0359 | 0.0037 | R | 0.0011 | 0.0065 |
| SCN × Tested × top-y × top-s | −0.000400 | R | 0.0026 | 0.0205 | R | −0.0019 | 0.0213 |
| SCN × Tested × top-y | 0.139*** | R | −0.0164 | 0.1633 | R | −0.0087 | 0.1667 |
| SCN × Tested × top-s | −0.00222 | R | 0.0016 | 0.0169 | R | 0.0051 | 0.0180 |
| SCN × Tested | −0.0863*** | R | −0.0054 | 0.0115 | R | −0.0021 | 0.0116 |
| Tested × top-y | −0.102** | R | 0.0104 | 0.1636 | R | 0.0085 | 0.1663 |
| Tested | 0.119*** | R | 0.0080 | 0.0092 | R | 0.0007 | 0.0103 |
| GT trait | 0.0924** | | 0.1023 | 0.0013 | | 0.1074 | 0.0014 |
| Nesting corr. | 0.810*** | | 0.8087 | 0.0008 | | 0.8106 | 0.0008 |
| Number of randomizations | — | | 1000 | — | | 1000 | — |

*Note*: Randomized variables are denoted by "R." All models include product age fixed effects, year fixed effects, CRD fixed effects, and brand fixed effects, as well as the use of three instrumental variables used in the baseline.
**p* < 0.1, ***p* < 0.05, ****p* < 0.01.

that, again, all of the coefficients associated with placebo treatments are statistically not different from zero. This lends support to our conclusion that the econometric results reported in Table 3 are indeed indicating a structural effect of the importance to farmers of the true SCN-related characteristics of soybean seed varieties.

## 6 | WELFARE

The analysis so far has focused on documenting the statistical significance of SCN-resistance and the ISU-SCN trial information for seed demand. Having estimated the discrete choice seed demand model, however, we are in a position to investigate the quantitative importance of these attributes. A first look at this question was provided by farmers' WTP estimates in Table 6. These WTP estimates can be combined with observed market data to provide a first-order approximation to the total surplus created by SCN resistance and ISU-SCN program information. To be specific, multiplication of WTP for a product characteristic by the corresponding acres planted to varieties that possess it gives an approximate welfare measure for that characteristic. Over the six years of the study (2011–2016) in Iowa and northern Illinois, we find that this method implies a total surplus change—from the combined availability of SCN-resistant varieties and the information associated with the ISU-SCN program—of about $466 million.

Whereas such a calculation is attractive because it relies simply on observed planted acres of each variety, in addition to the estimated WTPs, some limitations of this procedure are evident. One obvious drawback is the (rather dubious) presumption that the extent to which each variety is planted is independent of the attributes being evaluated. Furthermore, beyond total welfare increase, we also wish to evaluate the distribution of welfare gains. These limitations can be addressed by exploiting the structure of the estimated seed demand model.

A property of the nested logit model used in this study is that it permits a closed-form representation of welfare measures using the so-called log-sum formulae (e.g., Train, 2009). Comparing the expected value to the farmer of a choice set with a counterfactual situation where some attributes of the available choices are modified provides a structural approach to measuring welfare change due to the SCN resistance trait and ISU-SCN program. The two main counterfactual scenarios of interest are: (a) the case in which both the SCN resistance trait and the ISU-SCN program did not exist, and (b) the case in which only the ISU-SCN did not exist but SCN varieties were available. Comparing scenario (a) with the estimated baseline we can impute an overall welfare impact to the joint availability of varieties that are SCN resistant, and the distinct ISU-SCN extension program that has produced and diffused a copious amount of SCN-related information. Similarly, comparing scenario (b) with the observed baseline permits us to impute a value associated with the presence of ISU-SCN program information.

Relatedly, comparing the two counterfactual scenarios (a) and (b) permits us to identify the stand-alone value attributable to the availability of SCN-resistance varieties. The temptation is to assume that this value fully identifies the contribution of seed companies whose research and breeding efforts have made possible the development of such resistant varieties. Conversely, one would then associate the value from scenario (b) (relative to the baseline) as the welfare contribution attributable to the ISU-SCN program. As argued by a reviewer, this procedure may possibly overestimate the contribution of the extension program. After all, from Table 3, it appears that a major source of value appears associated with the mere fact of a variety being "tested" by the ISU-SCN program, irrespective of the variety's performance in the field experiments. Exactly what elements this value captures may be a matter of debate. Perhaps farmers think of the inclusion of a variety in the tested set as a signal of other inherent valuable characteristics that are ultimately attributable to breeding companies. In such a case, absent the ISU-SCN program, it is conceivable that other ways of conveying such information would arise in the marketplace. Hence, attributing the corresponding value entirely to the ISU-SCN program may overestimate the actual contribution of extension.

To address the foregoing concerns, we analyze an additional counterfactual scenario: (c) the case where only the information concerning varieties' performances are not available. Because this scenario focuses narrowly on the quantitative performance in the ISU-SCN variety trials (in terms of yields and to end-of-season SCN-infestation, metrics conceivably of direct interest to farmers), then comparison of this scenario with the baseline uncovers a value that arguably should be ascribed directly to the extension program being evaluated. Whereas this is only a fraction of the value recovered from counterfactual (b), it does provide a credible lower bound for the overall contribution of the ISU-SCN extension program to welfare.

For the evaluation of seed demand in all counterfactual scenarios hypothesized, the key is to determine what soybean variety prices would have been observed in the counterfactual scenarios. The standard approach in the literature would be to use a model of competition to generate equilibrium prices, such as the common Bertrand-Nash model (Houde, 2012; Nevo, 2001; Petrin, 2002). This standard pricing model for differentiated products, however, is not appropriate in our setting because one of the main seed sellers (Monsanto) holds a monopoly in the GT trait, which it licenses to its competitor. Whereas the terms of these licenses are not publicly known, the economics of licensing suggests that such arrangements may seriously affect firms' pricing choices and may indeed lead to a high rate of collusion (Shapiro, 1985).

To proceed, we rely on a reduced-form hedonic price approach, as in Hausman and Leonard (2002) and Ciliberto et al. (2019). The hedonic price function can be interpreted as a reduced-form approximation to the equilibrium prices in these differentiated-product markets (Pakes, 2003). For our model, the hedonic price regression is represented by

$$p_{jm} = \theta_0 + \sum\nolimits_{k=1}^{9} \theta_k Z_{jt[m]}^k + \phi x_j + \zeta_{jt[m]}^A + \zeta_{b[j]} + \zeta_{t[m]} + \zeta_{l[m]} + \sum\nolimits_{i=1}^{3} \rho_i W_{jm}^i + \mu_{jm} \tag{14}$$

where the $\zeta$ terms denote fixed effects; $W_{jm}^i$ denotes the instrumental variables discussed earlier; and, all other variables are defined as in Equation (9). Here, the parameter vector to be estimated is $(\zeta, \theta, \phi, \rho)$, and $\mu_{jm}$ is an error term.

It is apparent that the hedonic price equation in (14) is the same as the first-stage price equation of the 2SLS procedure, the estimates of which are reported in Column (1) of Table 4. Relying on the hedonic price regression, we predict prices for the baseline situation and for each of the counterfactual scenarios discussed in the foregoing. The predicted prices in the baseline are simply the fitted values from the estimation of Equation (14) and are denoted $\widehat{p}_{jm}$. Counterfactual prices for scenario ($a$) are given by:

$$\widetilde{p}_{jm} = \widehat{\theta}_0 + \widehat{\phi} \cdot x_j + \widehat{\zeta}^A_{jt[m]} + \widehat{\zeta}_{b[j]} + \widehat{\zeta}_{t[m]} + \widehat{\zeta}_{l[m]} + \sum\nolimits_{i=1}^{3} \widehat{\rho}_i W^i_{jm} \tag{15}$$

where the overstruck hat symbol denotes the estimated parameters from the hedonic equation in (14). Note that here we drop all information variables about ISU-SCN as well as the term for the SCN resistance per se.

As for scenario ($b$), the counterfactual prices are:

$$\widetilde{p}_{jm} = \widehat{\theta}_0 + \widehat{\phi} x_j + \widehat{\theta}_1 Z^1_{jt[m]} + \widehat{\zeta}^A_{jt[m]} + \widehat{\zeta}_{b[j]} + \widehat{\zeta}_{t[m]} + \widehat{\zeta}_{l[m]} + \sum\nolimits_{i=1}^{3} \widehat{\rho}_i W^i_{jm} \tag{16}$$

where we have dropped all information variables about ISU-SCN but retain those concerning SCN resistance. Finally, for scenario ($c$) the counterfactual prices are written as:

$$\widetilde{p}_{jm} = \widehat{\theta}_0 + \widehat{\phi} x_j + \widehat{\theta}_1 Z^1_{jt[m]} + \widehat{\theta}_5 Z^5_{jt[m]} + \widehat{\theta}_9 Z^9_{jt[m]} + \widehat{\zeta}^A_{jt[m]} + \widehat{\zeta}_{b[j]} + \widehat{\zeta}_{t[m]} + \widehat{\zeta}_{l[m]} + \sum\nolimits_{i=1}^{3} \widehat{\rho}_i W^i_{jm} \tag{17}$$

For each scenario, these predicted prices are used along with the estimated demand model of Equation (9) to obtain the *predicted* mean expected per-acre profit of product $j$ in market $m$, given any pair of $j$ and $m$ for each case—namely, $\widehat{\delta}_{jm}$ for the baseline and $\widetilde{\delta}_{jm}$ for the counterfactual scenario. Given such predicted mean expected profits, we calculate the aggregate value of the nests in the demand system—the "inclusive values" for the two situations being compared. The inclusive values in our one-level nesting structure are defined as follows (see Björnerstedt and Verboven (2016) for a more general setup):

$$I_{gm} = (1-\sigma) \cdot \ln \sum_{j \in J^1_m} \exp\left(\frac{\delta_{jm}}{1-\sigma}\right) \tag{18}$$

$$I_m = \ln\left(1 + \exp\left(I_{gm}\right)\right) \tag{19}$$

where $I_{gm}$ is the inclusive value of all soybeans (the nest for the inside option) in market $m$, and $I_m$ is the inclusive value for the entire choice set (including all soybeans and the outside option). By inserting $\widehat{\delta}_{jm}$ and $\widetilde{\delta}_{jm}$, respectively, into Equation (18), we can measure two predicted inclusive values for every market: $\widehat{I}_m$ (the value in the baseline) and $\widetilde{I}_m$ (the value in the counterfactual).

Inclusive values are directly related to the expected value of the maximum of the given set of choices, which has a closed-form representation when the unobserved random terms have the TIEV distribution (Anderson et al., 1992). Thus, the net change in inclusive value represents farmers' welfare change between the two scenarios, which can be converted to dollar terms by dividing by the (negative of the) price coefficient. Namely, the per-acre farmers' welfare change between the two scenarios is: $\Omega_m \equiv \left(\widetilde{I}_m - \widehat{I}_m\right) / -\beta$. Farmers' surplus in market $m$ is therefore computed as $\Omega_m \times M_m$, where $M_m$ is the "market size" of market $m$ (i.e., the total number of acres available for planting either corn or soybeans). Similarly, the total surplus change of farmers is $\sum_m \Omega_m \times M_m$.

It is important to note that, in these counterfactual situations, the assumption is that the functionality of seeds is not affected by dropping the SCN-related attributes. In other words, we are

TABLE 8 Welfare estimates, Iowa and northern Illinois, 2011–2016 ($ million)

| | SCN resistance & ISU-SCN | ISU-SCN extension | ISU-SCN performance scores only |
|---|---|---|---|
| Farmers' welfare gains | 324.77 | 205.46 | 86.45 |
| Seed suppliers' revenue change | 153.71 | 85.96 | 18.53 |
| Total ex post welfare change | 478.48 | 291.42 | 104.98 |

*Note*: Entries in columns are computed from the counterfactual procedures (*a*), (*b*), and (*c*) described in the text. All estimates are in 2011 dollars.

assuming that another comparable variety would have been developed and commercialized in lieu of the observed SCN resistant variety, implying that the number of elements in the choice set is the same across counterfactual scenarios. An alternative to such "keep all" procedure would be to simply drop SCN-resistant varieties from the choice set. The premise of this alternative procedure, therefore, would be that no other variety would have been developed instead of the SCN-resistant varieties. Thus, these two procedures represent two extreme situations that could have arisen absent efforts to develop SCN-resistant varieties, neither of which is entirely compelling. The main point to appreciate is that, in this discrete choice framework, the diversification effect brought about by an expanded choice set has value, per se, to the decision maker. This is because an expanded choice set for farmers increases the likelihood of a choice that better matches their growing conditions. That is, the expected utility from a logit model increases with the number of elements in the choice set, as noted by Petrin (2002). Because the "keep all" procedure is the most conservative, in the sense that it produces the smallest estimated welfare impacts of SCN-resistant varieties, in what follows we present the results for this procedure only.

The results of the counterfactual welfare analysis are reported in Table 8. We find the total surplus to farmers from the combined availability of SCN-resistant varieties and the ISU-SCN program information is $324.77 million (over all markets considered; that is, for Iowa and northern Illinois for the six-year period 2011–2016).[15] Farmers' surplus directly connected with the ISU-SCN program is estimated at $205.46 million. More nuanced interpretations of these estimated surplus gains can be obtained by asking what the return to farmers is, as predicted by the model, from the availability of performance indicators for all the ISU-SCN tested varieties. This estimate turns out to amount to $86.45 million, again over all markets considered. As discussed earlier, the value to farmers attributable to the existence of the ISU-SCN extension program, therefore, likely lies between the upper bound of $205.46 million and the lower bound of $86.45 million. Conceptually, the difference between the overall surplus to farmers due to the joint availability of SCN-resistant varieties and extension information, and the value specifically attributed to the ISU-SCN program, provides an estimate of the contribution of the development of SCN-resistant varieties to farmers' surplus. Based on the foregoing, therefore, the value of R&D and breeders' contributions to farmers' surplus is estimated to lie between $324.77 − $205.46 = $119.31 million and $324.77 − $86.45 = $238.32 million.

The predicted per-acre mean expected profits under the baseline and a counterfactual, namely $\widehat{\delta}_{jm}$ and $\widetilde{\delta}_{jm}$, can also be used to estimate seed suppliers' revenue changes. Specifically, these terms, along with the structural nested logit share equations, permit estimation of the shares $\widehat{s}_{jm}$ (in the baseline) and $\widetilde{s}_{jm}$ (in the counterfactual scenario), for every product $j$ and market $m$. Brand $b$'s revenue change in the counterfactual being evaluated can thus be written as $\Delta R_{bm} \equiv M_m \left[\sum_{j \in b} \widetilde{s}_{jm} \widetilde{p}_{jm} - \sum_{j \in b} \widehat{s}_{jm} \widehat{p}_{jm}\right]$. Total revenue change of the entire soybean seed industry,

[15]Because prices were deflated by a price index prior to estimating the seed demand model, as noted earlier, all monetary values in this section are expressed in 2011 dollars.

over all markets, is thus $\sum_m \sum_b \Delta R_{bm}$, which, as reported in Table 8, is estimated to amount to \$153.71 million (combined availability of SCN-resistant varieties and ISU-SCN program information). These increases in seeds suppliers' revenue reflect both the enhanced demand for SCN-resistant varieties, and the price premia that such varieties command in the baseline (relative to counterfactual scenarios).

Summing the farmers' welfare gains and the seed suppliers enhanced revenues in Table 8 yields an estimate of the (ex post) total returns to innovation. Overall, over the region and period of study, this welfare gain is \$478.48 million.[16] It should be emphasized that the foregoing estimated surpluses attributable to the innovation of SCN-resistant varieties, and the associated ISU-SCN extension program, have a different interpretation for farmers and seed sellers. The surplus captured by farmers is a true welfare gain due to innovation and extension information; that is, it represents additional expected profit, net of any imputed extra cost, which would have been realized otherwise. The fact that seed sellers' revenue is increased by their ability to offer SCN-resistant varieties, however, is an *ex post* return, best interpreted as the payoff of (costly) the R&D program necessary to develop and commercialize these varieties.

## 7 | CONCLUSION

In this paper we provide direct evidence on the value of innovation and associated extension information. In particular, we have studied the impact of SCN-resistant soybean varieties and the information produced by the ISU-SCN program. This study focuses on the time and region in which the dissemination of the relevant extension information has been greatest: Iowa and northern Illinois from 2011 to 2016. The empirical analysis is rooted in a discrete choice model of farmers' seed demand. Specifically, we estimate a one-level nested logit model. Because of the nature of the question addressed, the seed demand model was specified and estimated at the individual variety level. To the best of our knowledge, this is the first seed demand model formulated and estimated at this extremely refined level.

We find significant seed demand effects associated with variables coding for varieties' attributes associated with SCN resistance and SCN-related extension information. The estimated model also provides the vehicle to assess the welfare consequences of an innovation and a high-profile associated extension education program that has targeted the soybean cyst nematode, the most harmful soybean pathogen in North America. Specifically, we estimate the WTP of farmers for SCN-resistant varieties, and the separate extension program devoted to educating farmers about SCN, as well as developing and disseminating information pertaining to the performance of SCN-resistant varieties. Farmers' ceteris paribus WTP for SCN resistance varieties is estimated at \$2.69/acre. The WTP for a variety that is SCN resistant, that is included ISU-SCN variety trials, and that is a top performer (according to the two metrics used in the model) is \$9.62/acre, comparable in magnitude to farmers' WTP for the GT trait (about 18% of the average seed cost).

A fuller characterization of the welfare implications of the SCN-resistance innovation, and the related extension program that we have studied, is provided by counterfactual analyses that rely on the structure of the estimated demand model. Two main counterfactual scenarios are considered: the absence of both the SCN resistance trait and ISU-SCN, and the absence of the ISU-SCN program only. In addition, we also consider a scenario where only the performance metrics produced by the ISU-SCN program are absent. A conservative procedure to implement these counterfactuals suggests large farmers' gains from the joint availability of SCN-resistance varieties and the ISU-SCN program: a total of \$324.77 million (over all markets spanning Iowa and northern Illinois from 2011 to 2016). Depending on the interpretation one puts on the counterfactuals evaluated, the portion of this surplus gain specifically attributable to the ISU-SCN extension program ranges between \$86.45

[16]The regional distribution of welfare change is illustrated in Figure X4 in the online supplementary appendix.

million and $205.46 million. The remainder of the total farmers' surplus, ranging from $119.31 million and $238.32 million, represents the contribution attributable to the R&D and breeding activities responsible for the actual development of SCN-resistant varieties.

These counterfactual scenarios also provide additional insights into the distribution of these welfare gains. In particular, the seed industry has benefited from both the introduction of SCN-resistant varieties and also from the ISU-SCN program by an increase in soybean seed revenues (over the region and time frame of the study) estimated at $153.71 million. Thus, the seed industry appears to have appropriated about one-third of the estimated total ex post surplus change. In any case, we should note that the returns to breeders and farmers have a different interpretation in this context. For seed companies, the estimates pertain to *ex post* returns to the past R&D investments that made possible the novel SCN-resistant varieties, whereas for farmers the estimated surplus gain is a true welfare gain.

The results of this study carry some general implications. The extension program we have studied is, quite clearly, a "success story," and one should be mindful of that when extrapolating lessons to the broader set of agricultural extension (or research) activities, some of which may end up as dry holes. Yet, what appears to have made this program a success are three main factors: (a) it addressed a quantitatively important issue, a pathogen that can cause major yield losses to US farmers; (b) it involved a sustained extension effort spanning more than two decades; and, (c) it focused on the provision of a quintessential public good—the production and dissemination of information about the effectiveness of SCN-resistant varieties. These elements appear to be almost textbook checkboxes on how to prioritize extension activities and arguably may have broader application.

The fact that the welfare gains uncovered in this study ultimately rely on efforts exerted both by seed companies, who developed the SCN-resistant varieties, and by the ISU-SCN program, who provided third-party testing of their effectiveness, and dissemination of the associated information, is also noteworthy. This points to a strong complementarity between research and extension activities, a traditional justification for a good portion of land grant university work on agriculture. In an age where the private sector is undertaking an increasingly larger share of agricultural research, our results underscore the benefits of enabling institutional arrangements conducive to exploiting such synergy.

## ACKNOWLEDGMENTS

The authors are grateful to Dr. Greg Tylka, Iowa State University, for generously sharing his depth of expertise concerning SCN-resistant soybeans. The authors also thank editor Terrance Hurley and the journal's reviewers for their helpful comments. This project was supported in part by a USDA NIFA grant, contract number 2018-67023-27682.

## REFERENCES

Allen, Tom W., Carl A. Bradley, Adam J. Sisson, Emmanuel Byamukama, Martin I. Chilvers, Cliff M. Coker, Alyssa A. Collins, et al. 2017. "Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2010 to 2014." *Plant Health Progress* 18: 19–27.

An, Yonghong, and Wei Zhao. 2019. "Dynamic Efficiencies of the 1997 Boeing-McDonnell Douglas Merger." *Rand Journal of Economics* 50(3): 666–94.

Anderson, Jock R., and Gershon Feder. 2007. "Agricultural Extension." *Handbook of Agricultural Economics* 3: 2343–78.

Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. Cambridge: MIT Press.

Bandara, Ananda Y., Dilooshi K. Weerasooriya, Carl A. Bradley, Tom W. Allen, and Paul D. Esker. 2020. "Dissecting the Economic Impact of Soybean Diseases in the United States over Two Decades." *PLoS One* 15(4): e0231141.

Berry, Cristopher R., and Anthony Fowler. 2021. "Leadership or Luck? Randomization Inference for Leader Effects in Politics, Business, and Sports." *Science Advances* 7(4): eabe3404.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica: Journal of the Econometric Society* 36(4): 841–90.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *Rand Journal of Economics* 25(2): 242–62.

Bissonnette, Kaitlyn M., and Gregory L. Tylka. 2017. *Seed Treatments for Soybean Cyst Nematode*. Ames, IA: Iowa State University Extension and Outreach, *CROP 3142 Flyer*, August.

Björnerstedt, J., and F. Verboven. 2016. "Does Merger Simulation Work? Evidence from the Swedish Analgesics Market." *American Economic Journal: Applied Economics* 8: 125–64.

Bresnahan, Timothy F. 1987. "Competition and Collusion in the American Automobile Industry: The 1955 Price War." *Journal of Industrial Economics* 35(4): 457–82.

Cardell, N. Scott. 1997. "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory* 13: 185–213.

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

Debreu, Gerard. 1960. "Review of RD Luce, Individual Choice Behavior: A Theoretical Analysis." *American Economic Review* 50(1): 186–8.

Dinar, Ariel, Giannis Karagiannis, and Vangelis Tzouvelekas. 2007. "Evaluating the Impact of Agricultural Extension on Farms' Performance in Crete: A Nonneutral Stochastic Frontier Approach." *Agricultural Economics* 36: 135–46.

Evanson, Robert. 1997. "The Economic Contributions of Agricultural Extension to Agricultural and Rural Development." In *Improving Agricultural Extension*, edited by B. Swanson, R. Bentz, and A. Sofranko, 27–36. Rome: FAO.

Fernandez-Cornejo, Jorge, Seth Wechsler, Mike Livingston, and Lorraine Mitchell. 2014. *Genetically Engineered Crops in the United States*." Economic Research Report No. 162. Washington, DC: U.S. Department of Agriculture Econoimc Research Service.

Genius, Margarita, Phoebe Koundouri, Celine Nauges, and Vangelis Tzouvelekas. 2013. "Information Transmission in Irrigation Technology Adoption and Diffusion: Social Learning, Extension Services, and Spatial Effects." *American Journal of Agricultural Economics* 96: 328–44.

Hausman, Jerry A., and Gregory K. Leonard. 2002. "The Competitive Effects of a New Product Introduction: A Case Study." *Journal of Industrial Economics* 50(3): 237–63.

Hendricks, Nathan P., Aaron Smith, and Daniel A. Sumner. 2014. "Crop Supply Dynamics and the Illusion of Partial Adjustment." *American Journal of Agricultural Economics* 96(5): 1469–91.

Ho, Daniel E., and Kosuke Imai. 2006. "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 California Recall Election." *Journal of the American Statistical Association* 101(475): 888–900.

Houde, Jean-François. 2012. "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline." *American Economic Review* 102(5): 2147–82.

Jin, Yu, and Wallace E. Huffman. 2016. "Measuring Public Agricultural Research and Extension and Estimating their Impacts on Agricultural Productivity: New Insights from US Evidence." *Agricultural Economics* 47: 15–31.

Kim, Hyunseok, and GianCarlo Moschini. 2018. "The Dynamics of Supply: US Corn and Soybeans in the Biofuel Era." *Land Economics* 94(4): 593–613.

Koenning, Stephen R., and J. Allen Wrather. 2010. "Suppression of Soybean Yield Potential in the Continental United States by Plant Diseases from 2006 to 2009." *Plant Health Progress* 11(1): 5.

Kovo, Assaf, and Alon Eizenberg. 2017. *Inferring Market Definitions and Competition Groups from Empirically-Estimated Demand Systems: A Practitioner's Guide*. Jerusalem, Israel: Department of Economics Working Paper, Hebrew University of Jerusalem.

Ma, Xingliang, and Guanming Shi. 2015. "A Dynamic Adoption Model with Bayesian Learning: An Application to US Soybean Farmers." *Agricultural Economics* 46: 25–38.

Maffioli, Alessandro, Diego Ubfal, Gonzalo V. Baré, and Pedro Cerdán-Infantes. 2011. "Extension Services, Product Quality and Yields: The Case of Grapes in Argentina." *Agricultural Economics* 42: 727–34.

Magnier, Alexandre, Nicholas G. Kalaitzandonakes, and Douglas J. Miller. 2010. "Product Life Cycles and Innovation in the US Seed Corn Industry." *International Food and Agribusiness Management Review* 13(1030-2016-82866): 17.

Mourtzinis, Spyridon, and Shawn P. Conley. 2017. "Delineating Soybean Maturity Groups across the United States." *Agronomy Journal* 109(4): 1397–403.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69: 307–42.

Niblack, T.L. 2005. "Soybean Cyst Nematode Management Reconsidered." *Plant Disease* 89(10): 1020–6.

Pakes, Ariel. 2003. "A Reconsideration of Hedonic Price Indexes with an Application to PC's." *American Economic Review* 93(5): 1578–96.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110(4): 705–29.

Richards, Timothy J., and Celine Bonnet. 2018. "New Empirical Models in Consumer Demand." In *The Routledge Handbook of Agricultural Economics*, edited by Gail L. Cramer, Krishna P. Paudel, and Andrew Schmitz, 488–511. New York: Routledge.

Rotemberg, Martin. 2019. "Equilibrium Effects of Firm Subsidies." *American Economic Review* 109(10): 3475–513.

Sanderson, Eleanor, and Frank Windmeijer. 2016. "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables." *Journal of Econometrics* 190(2): 212–21.

Shapiro, Carl. 1985. "Patent Licensing and R&D Rivalry." *American Economic Review Papers and Proceedings* 75(2): 25–30.

Staton, Mike. 2013. *Sources of Information for Selecting Soybean Cyst Nematode-Resistant Varieties*. East Lansing, MI: Michigan State University Extension. December 18, 2013. Retrieved from. https://www.canr.msu.edu/news/sources_of_information_for_selecting_soybean_cyst_nematode_resistant_variet

Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by Donald W.K. Andrews and James H. Stock, 80–108, Cambridge University Press.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Tylka, Gregory L. 2012. *Soybean Cyst Nematode Field Guide*. Extension Publication CSI 12. Ames, IA: Iowa State University. Retrieved from. https://lib.dr.iastate.edu/extension_pubs/223

Tylka, Gregory L., and Cristopher C. Marett. 2021. "Known Distribution of the Soybean Cyst Nematode, Heterodera Glycines, in the United States and Canada in 2020." *Plant Health Progress* 22(1): 72–4.

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

Useche, Pilar, Barham L. Barham, and Jeremy D. Foltz. 2012. "Trait-Based Adoption Models Using ex-Ante and ex-Post Approaches." *American Journal of Agricultural Economics* 95: 332–8.

Wrather, J.A., T.R. Anderson, D.M. Arsyad, Y. Tan, L.D. Ploper, A. Porta-Puglia, H.H. Ram, and J.T. Yorinori. 2001. "Soybean Disease Loss Estimates for the Top Ten Soybean-Producing Countries in 1998." *Canadian Journal of Plant Pathology* 23(2): 115–2.

## 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:** Lee, Seungki, and GianCarlo Moschini. 2022. "On the value of innovation and extension information: SCN-resistant soybean varieties." *American Journal of Agricultural Economics* 1–26. https://doi.org/10.1111/ajae.12283