Estimating the Value of Innovation and Extension Information: SCN-Resistant Soybean Varieties

Seungki Lee, GianCarlo Moschini

Working Paper 20-WP 603
June 2020

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070
www.card.iastate.edu

Seungki Lee is PhD Student, Department of Economics, Iowa State University, Ames, IA 50010. E-mail: seungki@iastate.edu.

GianCarlo Moschini is Professor and Pioneer Chair in Science and Technology Policy, Department of Economics, Iowa State University, Ames, IA 50010. E-mail: moschini@iastate.edu.

For questions or comments about the contents of this paper, please contact GianCarlo Moschini, moschini@iastate.edu.

Estimating the Value of Innovation and Extension Information: SCN-Resistant Soybean Varieties

Seungki Lee and GianCarlo Moschini *

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* (ISU-SCN). We use two data sources: the ISU-SCN experimental data, and a rich proprietary dataset on farmers' seed purchases. Combining these data, we estimate the value of SCN-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 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 (WTP) for SCN-resistant varieties to be \$0.81/acre. Also, the additional WTP for extension information related to these varieties is about \$0.75/acre. Using counterfactual analyses, over the period and region of the study, the total surplus associated with the existence of, and information about, the SCN resistance trait is estimated to be about \$140 million. About two thirds of this surplus is captured by seed suppliers and one third is held by farmers.

Keywords: Agricultural Extension, Nested Logit, Resistant varieties, Seed Market, Soybean Cyst Nematode, Value of Information

JEL codes: Q12, Q16, O33, O14

* Seungki Lee is a PhD student and GianCarlo Moschini is Professor and Pioneer Chair in Science and Technology Policy, both in the Department of Economics, Iowa State University.
We thank Professor Greg Tylka, director of the ISU-SCN program, for his encouragement and helpful assistance throughout this project.

This research was supported in part by the National Institute of Food and Agriculture, U.S. Department of Agriculture, grant No. 2018-67023-27682.

1. Introduction

Agricultural extension has long been regarded as an important public service, especially as it relates to technology adoption (Evenson 1997; Anderson and Feder 2007). 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, Karagiannis, and Tzouvelekas 2007; Maffioli et al. 2011; Genius et al. 2013; Jin and Huffman 2016). 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 and 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 pest to soybean yields in North America (Wrather et al. 2001; Koenning and Wrather 2010; Allen et al. 2017). This plant parasite (a microscopic roundworm) feeds on soybean roots and can result in damages that have serious repercussion 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). 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. 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-distribute 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, along the lines of the framework developed by Ciliberto, Moschini, and Perry (2019). Given the nature of the research question addressed here, however, the seed demand model we specify is 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 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 Illinois together are composed of 18 CRDs). 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. Specifically, we use a one-level nested logit model, which maintains that, on a given plot, there is higher substitutability between soybean varieties 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 the effects of other potentially confounding factors, such as the product life cycle of commercial varieties.

Based on the estimated demand model, we separately 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 control within the test samples. Estimated WTPs provide a first-order approximation to the total surplus produced by the innovation (i.e., SCN resistance) and the distinct but related extension information. Also, using the observed price premia with respect to SCN resistance and extension information, we decompose the total surplus into net returns of farmers, and the change in revenue of seed suppliers. Our results from the demand estimation reveal that the WTP for the SCN resistance is $0.81/acre and WTP for being tested by ISU-SCN is $0.75/acre. Additionally, when a tested SCN-resistant variety performed above yield median, the extra WTP is found to be $1.36/acre. The total surplus attributable to the availability of SCN resistant varieties is calculated to be $62.52 million in Iowa and Illinois during 2011-2016, and the total surplus attributable to ISU-SCN is estimated at $65.98 million.

Additionally, a more structural welfare estimation is conducted by considering two counterfactual scenarios: (*i*) the absence of SCN resistance traits and the associated ISU-SCN extension program; and, (*ii*) the absence of the ISU-SCN program only. 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 $140.07 million of total welfare gains from the joint contribution of SCN resistance availability and the ISU-SCN extension program, with seed suppliers capturing 69% of this surplus and farmers obtaining 31%.

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. 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 (Wrather et al. 2001; Koenning and Wrather 2010; Allen et al. 2017). In the United States, the SCN was first discovered in North Carolina in 1954 and is currently found in more than 25 states (Tylka and Marett 2014, 2017). 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). 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 notoriously 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 the 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/100cc) at the end of the season, a measure of 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).[1] 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 each year have been documented. Importantly, since fall 2010, the reports have been directly mailed in the post-harvest season to (mainly) Iowa and northern Illinois farmers as a supplement to the magazine of *Iowa Farmer Today* and *Illinois Farmer Today*. Through these weekly periodicals the ISU-SCN reports have been distributed, free of charge, to more than 90,000 farm owners and operators in Iowa and Illinois.

---

[1] **Appendix B** provides a more detailed explanation about the ISU-SCN data.

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 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 soybeans 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 crop reporting district (CRD) level.[2] Kynetec data for the two states and period of interest (Iowa and Illinois over 2011-2016), include an average of 901 farmers per year and 2,734 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 has tested 1,904 varieties (1,798 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: 666 of the SCN-resistant tested varieties are observed in the seed data over the entire period; and, in the estimation period of 2011-2016, 417 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

---

[2] CRDs are aggregates of counties, as defined by National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA).

from various extension publications,[3] 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 is taken to represent the underlying raw value of innovation, which is brought about by breeding activities; and, we assume 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.

Figure 1 provides an overview of our data structure. In the six-year timeframe of this study, in Iowa and Illinois, we observe soybean seed purchases for 2,705 distinct varieties, among which 1,162 are SCN-resistant. Within these SCN-resistant varieties, 417 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 market share is accounted for by ISU-SCN tested resistant varieties in our study period, 2011-2016.

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 1** reports the nominal average prices for SCN-resistant and susceptible varieties, separately for conventional and glyphosate tolerant (GT) groups. While GT products

---

[3] 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 6,912 varieties.

are generally about $10/acre more expensive 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 $1.43/acre for conventional varieties and $0.74/acre for GT varieties.

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 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, Barham, and Foltz (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, Barham, and Foltz (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, as in Ciliberto, Moschini, and Perry (2019), we develop a nested logit model to improve on 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, Levinsohn, and Pakes (1995). Specifically, the regional level of our analysis, as in Ciliberto, Moschini, and Perry (2019), is the CRD. The area of our study, Iowa and Illinois, contains 18 CRDs (nine in each state) and the estimation period encompasses six years, from 2011 to 2016, as discussed earlier (thus, our analysis covers 108 markets).

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. These are by far the two most important crops in Iowa and Illinois, and this total area has remained very stable over the study period.

3.2. Brands

The supply structure of each soybean seed is specified by three levels: parent company, brand, and variety. There are five predominant companies in the U.S. soybean market. We categorize all seed products into six groups which are the big five companies, plus local and regional companies. A brand is a sub-company of the parent company and 106 brands are observed in Illinois and Iowa during 2011-2016. **Table 2** shows the average annual market shares for each brand that takes above 1% share. Brands with less than 1% share are defined by 'company name-others.' For example, 'Local & Regional Companies-others' indicates products that do not belong to the big five companies and have less than 1% average annual market share. As a result, we use 18 brand variables. From Table 2, we can see the relations between major parent companies and brands. In particular, the two most prominent brands from Dupont and Monsanto (Pioneer and Asgrow Seed Company, respectively) account for more than half of the market share. This market configuration is that of a concentrated, differentiated-product industry. From the perspective of demand estimation, equilibrium pricing in such industries inevitably raises the issue of potentially endogenous prices, which we will address at the estimation stage.

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, Moschini, and Perry (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 2,705 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 3 shows the average number of varieties in each market over the period of study, separately for conventional and GT varieties. It is apparent that the choice sets in our model are quite large, including an average of about 96 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.

3.4. Product Life Cycle

Because our analysis 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, Moschini, and Perry (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, Kalaitzandonakes, and Miller 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 32 states (not just Iowa and Illinois) and 21 years.[4] For varieties observed chosen by Iowa and Illinois farmers in our sample, the average Age is 3.29 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. **Figure 3** illustrates the average market share of all soybean varieties planted in Iowa and Illinois by their Age, for the period 2011-2016. It is apparent that a large portion of the market is accounted for by younger varieties, and that the popularity of varieties decreases considerably after their first 3-5 year of commercialization.

4. Seed Demand Model

Given a plot i, in market m, the farmer is assumed to choose the seed variety j that provides the largest expected profit. Ciliberto, Moschini, and Perry (2019) show that two reasonable properties of the per-acre production technology are particularly useful—constant returns to scale, and fixed proportion between land and seed. Given that, the per-acre expected profit from planting variety j in plot i can be expressed as:

$$(1) \qquad \pi_{ijm} = \pi_{ijm}\left(r, w; \mathbb{I}_j\right) - p_{jm}$$

where r is the (deflated) expected output price, w is the vector of (deflated) prices of all inputs other than seeds and land, and p_j is the (deflated) price of seed variety j (where price is expressed on a per-acre basis).[5] Note that the per-acre profit function $\pi_{ijm}\left(r, w; \mathbb{I}_j\right)$ is specified to depend on both the variety being planted as well as the plot. In particular, this profit function will depend on the attributes of variety j (including the presence of the GT trait, and the variety's SCN resistance attributes). To emphasize this point, the function is conditional on $\mathbb{I}_j$,

---

[4] 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 likely insignificant for the period 2011-2016 used in the econometric analysis.

[5] In the empirical application, nominal prices are deflated by the crop sector index for price paid, which is provided by USDA (index = 1 in 2011).

the set of relevant information pertaining to the variety. Note also that the linearity of the payoff function in seed prices in equation (1), which will be very convenient in the econometric specification below, is a direct consequence of the fixed proportion condition between seed and land.

Farmers' seed choice, in this context, can then be represented by a standard discrete-choice problem. Given that J_m varieties are available in market m, the farmer's profit-maximizing choice entails solving

$$\max_j \ \pi_{ijm}, \quad j \in \{0,1,\ldots, J_m\} \tag{2}$$

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

To make this model operational, we need to parameterize the per-acre profit function in equation (1). We represent this function in terms of observable and unobservable components as follows:

$$\begin{aligned} \pi_{ijm} &= \alpha \cdot p_{jm} + \gamma \cdot z_j + \sum_{k=1}^{4} \lambda_k A_{jt[m]}^{k} + \beta_1 \cdot x_j + \beta_2 \cdot \mathbb{I}_{jt[m]}^{\text{tested}} + \beta_3 \cdot \mathbb{I}_{jt[m]}^{\text{y-top50}} + \beta_4 \cdot \mathbb{I}_{jt[m]}^{\text{s-top50}} \\ &\quad + \xi_{t[m]} + \xi_{l[m]} + \xi_{b[j]} + \xi_{jm} + \varepsilon_{ijm} \end{aligned} \tag{3}$$

As per (1), the variety's price p_{jm} enters linearly in the profit equation (its coefficient here differs from unity because of the well-known unidentified scaling factor of the logit model); z_j is a dummy variable that codes for whether or not the variety in question includes the GT trait; and, the terms A_{jt}^{k} are dummy variables that code for the phase of commercial age of variety j in year t. Specifically, based on **Figure 3,** we make use of four age dummy variables corresponding, respectively, to varieties having commercial age 1-3, 4-6, 7-9, and 10 years or greater. [6] The next set of variables reflects information related to SCN resistance. The dummy variable x_j takes the value of 1 if this is SCN-resistant variety. $\mathbb{I}_{jt[m]}^{\text{tested}}$ is an indicator function

---

[6] Two other specifications for the age variable are tried for robustness checks, yielding virtually identical results (See **Appendix C**).

that takes the value of 1 if the variety in question has been tested by ISU-SCN by the year that defines market m, and 0 otherwise. The other two information variables that we include are indicator variables that code for the performance of a tested variety in the ISU-SCN trials. Specifically, $\mathbb{I}_{jt[m]}^{\text{y-top50}}$ is an indicator variable that takes the value of 1 if the variety in question was tested by ISU-SCN and performed better than the 50th percentile in terms of yield (value of 0 otherwise); and, $\mathbb{I}_{jt[m]}^{\text{s-top50}}$ is an indicator variable that takes the value of 1 if the variety was tested by ISU-SCN and performed better than the 50th percentile in terms of end-of-season SCN population density (a metric of SCN resistance), and value of 0 otherwise.[7]

The remaining terms in equation (3) 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, $\xi_{t[m]}$, $\xi_{l[m]}$, and $\xi_{b[j]}$ represent the time (year), region (CRD), and brand fixed effects; the term ξ_{jm} captures all other unobserved product-market specific components; and, the term ε_{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 α, $\gamma, \beta_1, \beta_2, \beta_3, \beta_4$, and the set of λ_k parameters (and, implicitly, all of the included fixed effects, of course).

4.1. Nested Logit

To proceed, we need to make assumptions about the distribution of unobserved terms. Following Berry (1994), it is convenient to represent with δ_{jm} all terms in equation (3) that are common to all plots in the same market and planted with the same variety. That is, δ_{jm}

---

[7] The performance metrics are defined by an average of ratios of a specific performance (yield or end-of-season SCN density) to the mean performance of the control group (non-resistant varieties). See **Appendix B** for additional explanations. **Appendix C** shows some alternative specifications used for robustness checks.

represents the mean expected per-acre profit of variety j in market m. Hence, per-acre profit can be represented as:

$$(4) \qquad \pi_{ijm} = \delta_{jm} + \varepsilon_{ijm}$$

From the problem in equations (2)-(3), and the structure in equation (4), observing the selection of variety j, in a given choice situation, means that

$$(5) \qquad \varepsilon_{ijm} - \varepsilon_{ikm} \geq \delta_{km} - \delta_{jm} \quad \forall k \in \{0, 1, \ldots, J_m\}$$

Assuming that the unobserved terms ε_{ijm} are identically and independently drawn from a Type I Extreme Value (TIEV) distribution, then choice probabilities (or, equivalently, market shares) take the familiar logit structure (Train, 2009):

$$(6) \qquad s_{jm} \equiv \Pr\left(\pi_{ijm} \geq \pi_{ikm}, \forall k \neq j\right) = \frac{\exp\left(\delta_{jm}\right)}{\sum_k \exp\left(\delta_{km}\right)}.$$

However, it is known 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 that the farmer would be equally likely to choose the outside option (planting 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).[8] 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.

To implement this one-level nesting model, more structure is needed for the plot-specific unobserved component, ε_{ijm}. Now that individual choices are grouped by inside option and outside option, we denote these two exclusive groups by $g \in \{0,1\}$, where $g = 0$ means a

---

[8] Saved soybean seeds are also included in the outside option as they are not tradable goods in the market. About 1.68% of soybean farmland per market are observed using saved seeds.

farmer chooses the outside option while $g = 1$ indicates the inside option. Given that, the unobserved component is written as:

(7) $$\varepsilon_{ijm} = v_{igm} + \left(1-\sigma\right) v_{ijm}$$

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 $\sigma \in \left[0,1\right]$ captures correlation between varieties within the inside option group. The term v_{igm} is assumed to have the unique distribution such that ε_{ijm} again follows the TIEV distribution (Cardell, 1997). The larger the σ, the stronger the correlation between the varieties within the group. That is, if σ 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 σ approaches 0 the model reduces to the simple logit.

Similar to equation (6), we can derive 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

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

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 $\pi_{i0m} = \varepsilon_{i0m}$.[9] The probabilities of choosing the inside option (S_{1m}), and outside option (s_{0m}), are respectively

---

[9] As the mean value of outside option is a reference point, the value itself does not affect the comparison between alternatives.

$$S_{1m} = \frac{\left[\sum_{k\in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{(1-\sigma)}}{1+\left[\sum_{k\in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{(1-\sigma)}}, \text{ and} \tag{9}$$

$$s_{0m} = \frac{1}{1+\left[\sum_{k\in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{(1-\sigma)}} \tag{10}$$

Then, the unconditional probability of choosing a soybean variety j, defined as $s_{jm} \equiv s_{jm|g=1} \cdot S_{1m}$, is:

$$s_{jm} = \frac{\exp\left(\frac{\delta_{jm}}{1-\sigma}\right)}{\left[\sum_{k\in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{\sigma} \cdot \left\{1+\left[\sum_{k\in J_m^1} \exp\left(\frac{\delta_{km}}{1-\sigma}\right)\right]^{(1-\sigma)}\right\}}. \tag{11}$$

Using Berry's (1994) inversion, and recalling the definition of the mean profit terms δ_j, the log ratio of market shares is:

$$\begin{aligned} \ln\left(s_{jm}\right) - \ln\left(s_{0m}\right) = {} & \alpha \cdot p_{jm} + \gamma \cdot z_j + \beta_1 \cdot x_j + \beta_2 \cdot \mathbb{I}_{jt[m]}^{\text{tested}} + \beta_3 \cdot \mathbb{I}_{jt[m]}^{\text{y-top50}} + \beta_4 \cdot \mathbb{I}_{jt[m]}^{\text{s-top50}} \\ & + \sigma \ln\left(s_{jm|g=1}\right) + \sum_{k=1}^{4} \lambda_k A_{jt[m]}^{k} + \xi_{t[m]} + \xi_{l[m]} + \xi_{b[j]} + \xi_{jm} \end{aligned} \tag{12}$$

This is the linear regression equation that we estimate.

4.2. *Identification*

The key identification issue, in this setting, is related to the possible price endogeneity in equation (12) (and also the endogeneity of the conditional share appearing on the right-hand side of equation (12)). As shown in **Table 2**, 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 ξ_{jm} and p_{jm} in equation (12). 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 α 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, Levinsohn, and Pakes (1995). Specifically, we follow Ciliberto, Moschini, and Perry (2019) and use functions of the traits in competing varieties as our instrumental variables. 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.

Because this study focuses on soybeans, for which we have only one GT trait, we end up using six instrumental variables; the number of competing products by variety, brand, and parent company, interacted with two underlying trait configurations: conventional and GT varieties. Nevo (2000) points out that if there are small variations of the products offered in different markets, the instrumental variables cannot show a meaningful variation between markets because of the brand dummy variable. As shown in **Table 3**, our sample shows a good amount of variations in products between markets.

5. Estimation Results

Table 4 reports estimation results for the demand model. We report results from OLS estimation, for comparison purposes and to assess the effect of the instrumental variable procedure we implement.

Because most likely OLS estimates are plagued by price endogeneity, the price coefficient is biased and not significantly different than zero. Conversely, two stage least squared (2SLS) estimates rely on instrumental variables for the endogenous prices and conditional market shares, $\ln\left(s_{jm|g=1}\right)$. The need for the latter, which arises from the nested logit specification, is that $\ln\left(s_{jm|g=1}\right)$ would be correlated with ξ_{jm} if our dependent variable, $\ln\left(s_{jm}\right)-\ln\left(s_{0m}\right)$, had a correlation with ξ_{jm}. In the first stage regression, the aforementioned six instrumental variables are used. Column (2) of **Table 4** presents estimated parameters for the 2SLS. We note that, in this case, the estimated price coefficient has the expected negative sign, which suggests that the instrumental variables are effective. From here on, we focus on the 2SLS results.

It is of some interest to express our estimates in terms of elasticities. Specifically, the own-price elasticity (for $j \neq 0$) can be obtained by (for notational simplicity we drop the market subscript

$$\frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} = \alpha p_j \left[\frac{1}{1-\sigma} - \left(\frac{\sigma}{1-\sigma}\right)\frac{s_j}{s_0} - s_j \right].$$

Cross-price elasticity within the nest (inside options, soybeans) is calculated by

$$\frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = -\alpha p_k \left[\left(\frac{\sigma}{1-\sigma}\right)\frac{s_k}{s_0} - s_k \right] \quad \text{for } j \neq k \in \{1, 2, \ldots, J\}.$$

Cross-price elasticity across the nest (i.e., inside option and outside option) is calculated by

$$\frac{\partial s_0}{\partial p_j}\frac{p_j}{s_0} = -\alpha p_j s_j \quad \text{for } j \in \{1, 2, \ldots, J\}.$$

The own-price elasticity is found to be around -13.97, which indicates that individual seed demand functions are quite elastic. This is not surprising, considering that we are modeling demand at the (very fine) variety level. Cross-price elasticity within the nest is about 0.126, and cross-price elasticity across the nest is about 0.0001. This shows that, 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).

Table 4 also reports the estimated coefficients for the set of Age dummy variables included in the model. Seed varieties appear most preferred in the first three years of commercialization, and their appeal starts to show a sizeable decline, *ceteris paribus*, starting with their fourth year.

The decline is particularly strong after the second phase. Interestingly, the coefficient associated with commercial age greater than or equal to 10 years is less negative that that for ages 7 to 9 years. Apparently, whereas the likelihood of a variety surviving on the market beyond a few years is low, but varieties that succeed can be around for a long time.

6. Welfare

A first look at the welfare impact of SCN-related innovation and extension is provided by farmers' willingness to pay, which can be evinced by the estimated demand model. Alternatively, the structural nature of the model permits us to characterize alternative counterfactual scenarios that provide further insights into the value of innovation and extension information related to SCN.

6.1. Willingness to Pay and Surplus Calculation

Based on the demand estimation, we measure welfare changes attributable to the product characteristics of interest: SCN resistance *per se*, and the information produced by the ISU-SCN program. First of all, we can calculate the (marginal) WTP for SCN resistance by $-\beta_1/\alpha$ (Train, 2009). In a similar way, the WTP for information provided by ISU-SCN can be calculated by $-\beta_2/\alpha$, $-\beta_3/\alpha$, and $-\beta_4/\alpha$. From the estimates reported in **Table 4**, we find that farmers' WTP for SCN resistance is \$0.81/acre. The WTP for SCN-resistant seeds which are tested by ISU-SCN is \$0.75/acre additional to the WTP for SCN resistance, and when they performed above the median in terms of yield, within the test group, the extra WTP by the performance information is \$1.36/acre (performing above the median in terms of the SCN metric, by contrast, does not confer additional value). For a comparison purpose, we note that the model's estimates imply a WTP for the GT trait ($-\gamma/\alpha$) of \$9.30/acre.

These WTP estimates can be combined with quantity to measure the surplus change, which is the first-order approximation to the total surplus created by SCN resistance and ISU-SCN information. To be specific, the multiplication of WTP for information (or product characteristic) and corresponding acres planted by the seeds provides an approximate welfare measure for the effect of the ISU-SCN program. The caveat of this approach is to require a premise of which the number of acres planted by tested seeds (or SCN-resistant varieties) would have been the same even if they had not been tested by ISU-SCN. Because this information and SCN resistance do not affect agricultural practice (unlike the GT trait, for example), this assumption appears tenable to some degree.

Over the six years of the study (2011-2016), in Iowa and Illinois, we find that SCN resistance has brought about a total surplus of $62.52 million (on average, $0.58 million for each CRD in each year). Note that this value does not include the surplus of SCN resistance which is generated through the extension information. For the total surplus from ISU-SCN information, we calculate:

$$\left(WTP_{tested} \times acres_{tested} \right) + \left(WTP_{yield}^{top50} \times acres_{yield}^{top50} \right) + \left(WTP_{SCN}^{top50} \times acres_{SCN}^{top50} \right),$$

which is equal to $65.98 million.

Observed price premiums between SCN-resistant and other varieties provide a way to gauge how much surplus is captured by seed suppliers. This is useful as it helps us understand surplus distribution between farmers and seed suppliers. When computing the price differentials between SCN-resistant varieties and SCN-susceptible varieties, it is important to account for whether or not the variety is a conventional variety or a GT variety. Therefore, we use an average of price differences based on seed traits, weighted by total acres for each seed trait (i.e., GE or non-GE). The resulting price premium for SCN resistance, accounting for the seed traits, is $1.23/acre. A comparison of this number with the WTP for SCN resistance implies that seed suppliers essentially capture the entire total surplus from SCN resistance, and farmers' net returns are actually negatively affected.

For the price premium of having the test information (i.e., just being tested), we find that average per-acre price of the tested seeds is $0.97 higher than the untested SCN-resistant varieties. Within the tested, top 50% group in yield has an additional price premium of $1.36/acre, while the top 50% group in SCN control has - $0.81/acre price premium. This implies that yield information drives additional price premium while SCN information does not. Using the price premia, we calculate

$$\left(\text{P-premium}_{tested} \times acres_{tested}\right) + \left(\text{P-premium}_{yield}^{\text{top50}} \times acres_{yield}^{\text{top50}}\right) + \left(\text{P-premium}_{SCN}^{\text{top50}} \times acres_{SCN}^{\text{top50}}\right)$$

to approximate the change in firm revenues, which is \$65.10 million and 98.7% of the total surplus generated by ISU-SCN. **Table 7** provides the summary of surplus values. The largely skewed distribution of welfare might result from using the average of price differentials as price premium. The computed price premium does not fully control for critical factors to prices, such as product age or brand, and this can consequently affect the surplus calculation.

Whereas the calculations reported in the foregoing are attractive because they rely simply on observed price and quantity data, in addition to the estimated WTP, we should note some limitations of this procedure. One obvious drawback is the presumption that seed prices and quantities are unaffected by the introduction of SCN-resistant varieties whereas, in fact, one would expect that the availability of such varieties expands total soybean plantings, at the margin, and may put downward pressure on the prices of susceptible varieties. In principle, these limitations can be addressed by exploiting the structure of the estimated seed demand model.

6.2. Counterfactual Scenarios and Welfare Change Measures

Alternatively, we can use a more structural approach to compute the welfare change due to the SCN resistance trait and ISU-SCN program. The counterfactual scenarios we consider 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. For this kind of exercise, 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 Nash-Bertrand model (Nevo 2001, Petrin 2002, Houde 2012). As discussed in Ciliberto, Moschini, and Perry (2019), however, this structural assumption is problematic in our setting because of the existence of (unobserved) cross-licensing arrangements in the seed industry, which could distort seed suppliers' pricing decisions away from the standard Nash-Bertrand competition.

To circumvent this issue, we rely on a reduced-form hedonic price approach, as in Hausman and Leonard (2002) and Ciliberto, Moschini and Perry (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

$$(13)\quad \begin{aligned} p_{jm} &= \theta_1 \cdot x_j + \theta_2 \cdot \mathbb{I}^{\text{tested}}_{jt[m]} + \theta_3 \cdot \mathbb{I}^{\text{y-top50}}_{jt[m]} + \theta_4 \cdot \mathbb{I}^{\text{s-top50}}_{jt[m]} + \phi_0 \cdot z_j \\ &+ \zeta_{b[j]} + \zeta_{t[m]} + \zeta_{l[m]} + \sum_{k=1}^{4} \phi_k A^k_{jt[m]} + \mu_{jm} \end{aligned}$$

where the ζ terms denote fixed effects, and all other variables are defined as in equation (12). Here $\theta_1, \theta_2, \theta_3, \theta_4$, and ϕ_k for every k, are parameters to be estimated, and μ_{jm} is an error term.

Table 5 reports the estimated parameters of the hedonic price regression of equation (13). From this table, it seems yield information more directly affects prices than does SCN information. Relying on the hedonic price regression, we predict prices for three scenarios—the baseline situation, and the two aforementioned counterfactual scenarios. The predicted prices in the presence of ISU-SCN are given by:

$$(14)\quad \hat{p}_{jm} = \hat{\theta}_1 \cdot x_j + \hat{\theta}_2 \cdot \mathbb{I}^{\text{tested}}_{jt[m]} + \hat{\theta}_3 \cdot \mathbb{I}^{\text{y-top50}}_{jt[m]} + \hat{\theta}_4 \cdot \mathbb{I}^{\text{s-top50}}_{jt[m]} + \hat{\phi}_0 \cdot z_j + \hat{\zeta}_{b[j]} + \hat{\zeta}_{t[m]} + \hat{\zeta}_{l[m]} + \sum_{k=1}^{4} \hat{\phi}_k A^k_{jt[m]}$$

The counterfactual prices in the absence of SCN resistance are given by:

$$(15)\quad \tilde{p}_{jm} = \hat{\phi}_0 \cdot z_j + \hat{\zeta}_{b[j]} + \hat{\zeta}_{t[m]} + \hat{\zeta}_{l[m]} + \sum_{k=1}^{4} \hat{\phi}_k A^k_{jt[m]}$$

Note that we drop all information variables about ISU-SCN as well as the $\hat{\theta}_1 \cdot x_j$ term. This is because the extension information we study is predicated on this technology actually being available. As a result, the counterfactual analysis gives us aggregate welfare gains from both SCN resistance and ISU-SCN. The other counterfactual prices in the absence of ISU-SCN are:

$$(16)\quad \tilde{p}_{jm} = \hat{\theta}_1 \cdot x_j + \hat{\phi}_0 \cdot z_j + \hat{\zeta}_{b[j]} + \hat{\zeta}_{t[m]} + \hat{\zeta}_{l[m]} + \sum_{k=1}^{4} \hat{\phi}_k A^k_{jt[m]}$$

These predicted prices are plugged into the estimated demand model of equation (12). Since we have all other variable values and estimates for parameters, as shown in **Table 4**, we can obtain the *predicted* mean expected per-acre profit of product j in market m, given any pair of j and m for each case (i.e., $\hat{\delta}_{jm}$ for the presence of ISU-SCN and $\tilde{\delta}_{jm}$ for the absence of ISU-SCN).

Once predicted mean expected profits are obtained, we can calculate the aggregate value of the nests in the demand system, called "inclusive values" (Björnerstedt and Verboven 2016), for the two compared cases. The inclusive values in our one-level nesting structure are defined as follows:

$$(17) \qquad I_{gm} = (1-\sigma)\cdot \ln \sum_{j \in J_m^1} \exp\left(\frac{\delta_{jm}}{1-\sigma}\right)$$

$$(18) \qquad I_m = \ln\left(1 + \exp\left(I_{gm}\right)\right)$$

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 corn. By inserting $\hat{\delta}_{jm}$ and $\tilde{\delta}_{jm}$, respectively, into equation (17), we can measure two predicted inclusive values for every market: $\hat{I}_m$ and $\tilde{I}_m$. 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, De Palma, and Thisse 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 ($-\alpha$). Namely, per-acre farmers' welfare change from the ISU-SCN is:

$$(19) \qquad \Omega_m \equiv \frac{\left(\tilde{I}_m - \hat{I}_m\right)}{-\alpha}.$$

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 due to the ISU-SCN program 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 resistance attribute. In other words, we are 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. In what follows, we refer to this counterfactual as the "Keep all" scenario. Alternatively, we can simply drop SCN-resistant varieties from the choice set, which is labeled the "Naïve" scenario. Specifically for the "Naïve" counterfactual, when calculating $\tilde{I}_{gm}$ in equation (17), all $\tilde{\delta}_{jm}$ where j is SCN resistant (including all the tested by ISU-SCN) are excluded in the summation to reflect the loss of utility due to the unavailability of SCN resistance. The premise of the Naïve counterfactual, therefore, is that none of the SCN-resistant varieties would otherwise have existed. That is, this choice set rules out that other comparable varieties, without SCN resistance, could have been developed. Thus, these two scenarios represents two extreme configurations that could have arisen absent efforts to develop SCN-resistant varieties. The "Keep all" scenario provides the largest counterfactual choice set, whereas the "Naïve" scenario postulates the smallest choice set.

Neither scenario 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) in his study of the welfare effects of minivans. The "Keep all" scenario is the most conservative, in the sense that it produces the smallest estimated welfare impacts of SCN-resistant varieties. The "Naïve" scenario, conversely, because the choice set is severely restricted in the counterfactual, produces inflated welfare effects.

The results of the counterfactual welfare analysis are reported in **Table 6**. Based on the "Keep all" counterfactual, we find the surplus of farmers from ISU-SCN information is \$35.74 million, and the combined surplus to farmers due to both SCN-resistant varieties and ISU-SCN information is \$43.84 million (over the six years of analysis, for both Iowa and Illinois). The

"Naïve" counterfactual, by contrast, estimates the surplus of farmers from ISU-SCN information at \$193.44 million, with \$463.70 million as the combined welfare increase of farmers due to both SCN resistance trait and ISU-SCN information.

The predicted per-acre mean expected profits, $\hat{\delta}_{jm}$ and $\tilde{\delta}_{jm}$, can also be used to estimate the seed suppliers' revenue change. Specifically, by entering $\hat{\delta}_{jm}$ and $\tilde{\delta}_{jm}$, respectively, in equation (11), $\hat{s}_{jm}$ and $\tilde{s}_{jm}$ are obtained for every j and m. Brand b's revenue change due to the ISU-SCN can be written as $\Delta R_{bm} \equiv M_m \times \left(\sum_{j \in b} \left(\tilde{s}_{jm} - \hat{s}_{jm} \right) \right) \cdot \hat{p}_{jm}$. Total revenue change of the entire soybean seed industry in Iowa and Illinois for the period 2011-2016, that is $\sum_m \sum_b \Delta R_{bm}$, for the "Keep all" scenario is found to be \$43.06 million because of the availability of SCN-resistant varieties, and \$96.23 million due to both these varieties' availability and the ISU-SCN program. Hence, overall, these counterfactual analyses suggest that about two thirds of the total welfare gains are captured by seed suppliers, with farmers retaining about one third. The "Naïve" counterfactual, as expected, predicts much larger revenue changes (about four times those of the "Keep all" case).

It should be emphasized that 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.

Figure 4 exhibits the geographic distribution of our welfare estimates, by CRD in Iowa and Illinois, for the "Keep all" counterfactual analysis (the conservative scenario). Here we represent the total estimated welfare effects associated with both SCN-resistant varieties and the ISU-SCN extension program.[10] The patterns shown in **Figure 4** look plausible vis-à-vis the distribution of SCN pest pressure and the focus of the ISU-SCN extension program. In particular, we observe considerably larger welfare gains in the Iowa CRDs.

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 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.

The estimated model 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 pest 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. The WTP for SCN resistance is estimated at $0.81/acre. Also, the WTP for being tested by ISU-SCN is $0.75/acre, and when seeds performed above the median in terms of yield within the tested group, the additional WTP from the performance information is estimated at $1.36/acre. The first-order surplus approximation using the WTPs tells us that about $62.52 million of surplus are brought from SCN resistance, and $65.98 million of surplus are added by ISU-SCN.

---

[10] The welfare distribution from the counterfactual case for the absence of only ISU-SCN turns out to have a similar shape of Figure 4, only the scale becomes smaller.

Alternatively, combined with a hedonic price regression, the estimated demand model enables us to perform counterfactual analyses. We analyze two counterfactual scenarios: the absence of both the SCN resistance trait and ISU-SCN, and the absence of the ISU-SCN program only. The most conservative of the counterfactual analyses performed (using the "Keep all" choice set) produces welfare results similar to the approximation predicated on the direct use of WTP estimates: the total welfare gains from SCN resistance are $61.27 million, and the welfare gains from the ISU-SCN only are $78.80 million. Using the "Naïve" scenario, by contrast, produces much larger welfare estimates. Also, it seems that the seed industry has been able to appropriate about two thirds of the *ex post* returns to innovation and extension information.

The implications of this study are twofold. First, we provide direct evidence of the effectiveness of a targeted and well-defined agricultural extension program. As indicated by the estimated WTPs, the information produced and disseminated by extension is valuable to farmers. In the process, we also verify the value of the breeding activities that have made possible the realization of an array of SCN-resistant varieties. Second, we find that the distribution of the *ex post* welfare gains favors seed sellers, who capture approximately two thirds of the estimated gains. This is perhaps not surprising, and indeed such *ex post* increased seed revenues are best interpreted as returns to the past R&D investments that made possible the novel SCN-resistant varieties. It is also apparent, however, that seed sellers also benefit directly from the production and dissemination of SCN-related information. This point 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 in undertaking an increasingly larger share of agricultural research, enabling institutional arrangements conducive to exploiting such synergy may be warranted.

References

Allen, T. W., Bradley, C. A., Sisson, A. J., Byamukama, E., and others (43 authors in total). (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.

Anderson, S.P., De Palma, A. and Thisse, J.F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press.

Anderson, J. R., and Feder, G. (2007). Agricultural extension. *Handbook of Agricultural Economics, 3*, 2343-2378.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242-262.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.

Björnerstedt, J., and Verboven, F. (2016). Does merger simulation work? Evidence from the Swedish analgesics market. *American Economic Journal: Applied Economics, 8*, 125-64.

Bresnahan, T. F. (1987). Competition and collusion in the American automobile industry: The 1955 price war. *The Journal of Industrial Economics*, 457-482.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory, 13*, 185-213.

Ciliberto, F., Moschini, G., and Perry, E. (2019). Valuing product innovation: Genetically engineered varieties in US corn and soybeans. *The RAND Journal of Economics, 50(3), Fall 2019, 615-644.*.

Debreu, G. (1960). Review of RD Luce, Individual choice behavior: A theoretical analysis. *American Economic Review*, 50(1), 186-188.

Dinar, A., Karagiannis, G., and Tzouvelekas, V. (2007). Evaluating the impact of agricultural extension on farms' performance in Crete: A nonneutral stochastic frontier approach. *Agricultural Economics, 36*, 135-146.

Evanson, R. (1997). The economic contributions of agricultural extension to agricultural and rural development. In: Swanson, B., Bentz, R., and Sofranko, A., Eds. *Improving Agricultural Extension*, pp. 27-36. FAO, Rome.

Fernandez-Cornejo, J., Wechsler, S., Livingston, M., & Mitchell, L. (2014). *Genetically Engineered Crops in the United States*. Economic Research Report No. 162, USDA-ERS.

Genius, M., Koundouri, P., Nauges, C., and Tzouvelekas, V. (2013). Information transmission in irrigation technology adoption and diffusion: Social learning, extension services, and spatial effects. *American Journal of Agricultural Economics, 96*, 328-344.

Hausman, J. A., and Leonard, G. K. (2002). The competitive effects of a new product introduction: A case study. *The Journal of Industrial Economics*, *50*(3), 237-263.

Houde, J. F. (2012). Spatial differentiation and vertical mergers in retail markets for gasoline. *American Economic Review*, *102*(5), 2147-82.

Jin, Y., and Huffman, W. E. (2016). Measuring public agricultural research and extension and estimating their impacts on agricultural productivity: New insights from US evidence. *Agricultural Economics*, *47*, 15-31.

Koenning, S. R., and Wrather, J. A. (2010). Suppression of soybean yield potential in the continental United States by plant diseases from 2006 to 2009. *Plant Health Progress*, *11*(1), 5.

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

Maffioli, A., Ubfal, D., Baré, G. V., and Cerdán-Infantes, P. (2011). Extension services, product quality and yields: The case of grapes in Argentina. *Agricultural Economics*, *42*, 727-734.

Magnier, A., Kalaitzandonakes, N. G., and Miller, D. J. (2010). Product life cycles and innovation in the US seed corn industry. *International Food and Agribusiness Management Review*, *13*(1030-2016-82866), 17.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy*, *9*, 513-548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, *69*, 307-342.

Niblack, T. L. (2005). Soybean cyst nematode management reconsidered. *Plant Disease*, *89*(10), 1020-1026.

Pakes, A. (2003). A Reconsideration of Hedonic Price Indexes with an Application to PC's. *American Economic Review*, 93(5), 1578-1596.

Richards, T. J., and Bonnet, C. (2018). New empirical models in consumer demand. In: *The Routledge Handbook of Agricultural Economics* (pp. 488-511). Routledge.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, *110*(4), 705-729.

Staton, M. *Sources of information for selecting soybean cyst nematode-resistant varieties*. MSU Extension. December 18, 2013. Retrived from https://www.canr.msu.edu/news/sources_of_information_for_selecting_soybean_cyst_nematode_resistant_variet

Train, K. E. (2009). *Discrete Choice Methods with Simulation.* Cambridge University Press.

Tylka, G. L. (2012). Soybean Cyst Nematode Field Guide. *Extension and Outreach Publications, 223*. Retrieved from https://lib.dr.iastate.edu/extension_pubs/223

Tylka, G. L., and Marett, C. C. (2014). Distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada: 1954 to 2014. *Plant Health Progress*, *15*(2), 85-87.

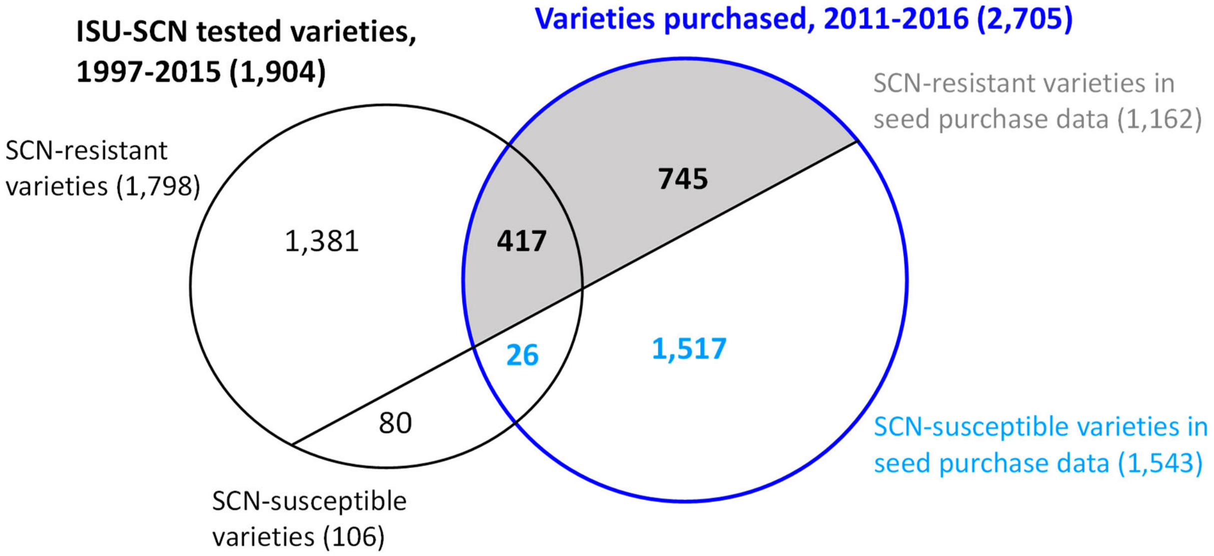
Tylka, G. L., and Marett, C. C. (2017). Known distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada, 1954 to 2017. *Plant Health Progress*, *18*(3), 167-168.

Useche, P., Barham, B. L., and Foltz, J. D. (2009). Integrating technology traits and producer heterogeneity: A mixed-multinomial model of genetically modified corn adoption. *American Journal of Agricultural Economics, 91*, 444-461.

Useche, P., Barham, B. L., and Foltz, J. D. (2012). Trait-based adoption models using ex-ante and ex-post approaches. *American Journal of Agricultural Economics, 95*, 332-338.

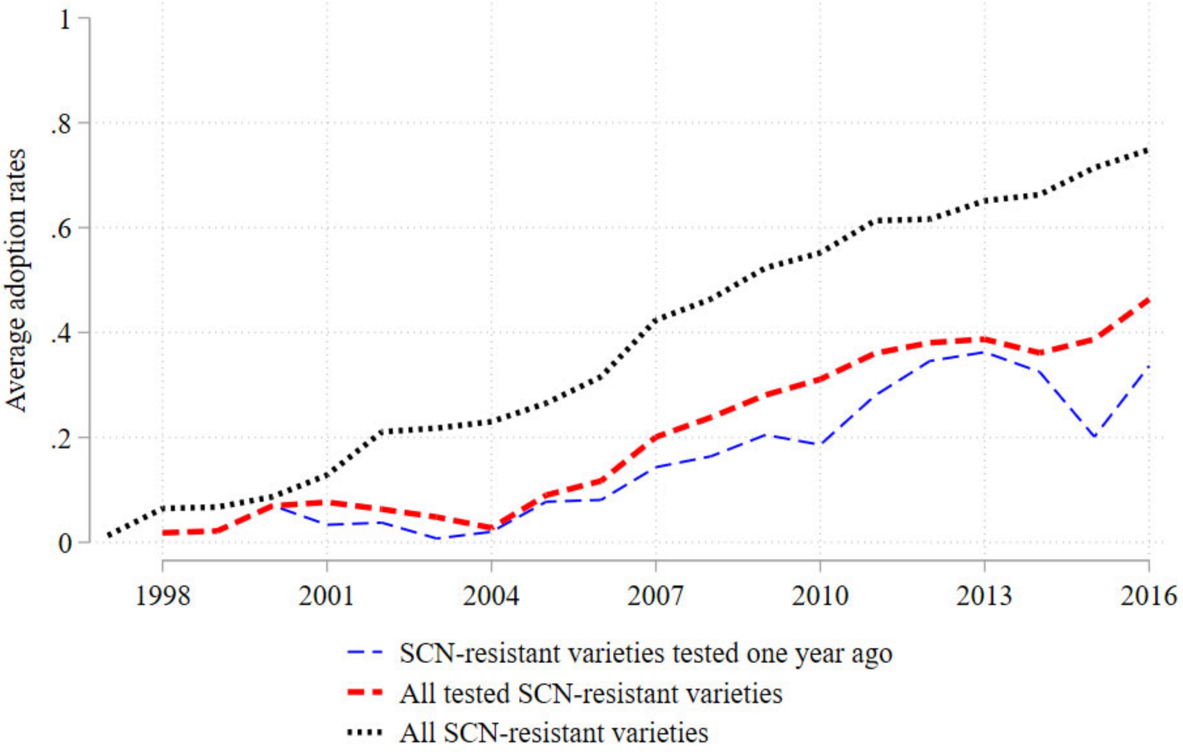
Wrather, J. A., Anderson, T. R., Arsyad, D. M., Tan, Y., Ploper, L. D., Porta-Puglia, A., Ram, H. H., and Yorinori, J. T. (2001). Soybean disease loss estimates for the top ten soybean-producing countries in 1998. *Canadian Journal of Plant Pathology*, *23*(2), 115-12.

Figure 1. Number of soybean seed varieties: ISU-SCN and Kynetec data, Iowa and Illinois



Note: 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.

Figure 2. Market share for SCN-resistant soybean seeds, Iowa and Illinois, 1997-2016



Note: Each line shows a trend of average market share of 18 CRDs for one of the three groups: (*a*) SCN-resistant varieties tested by ISU-SCN one year ago; (*b*) SCN-resistant varieties tested at some point in the past; and, (*c*) all SCN-resistant varieties.

Figure 3. **Average market share by soybean varieties' Age, Iowa and Illinois, 2011-2016**

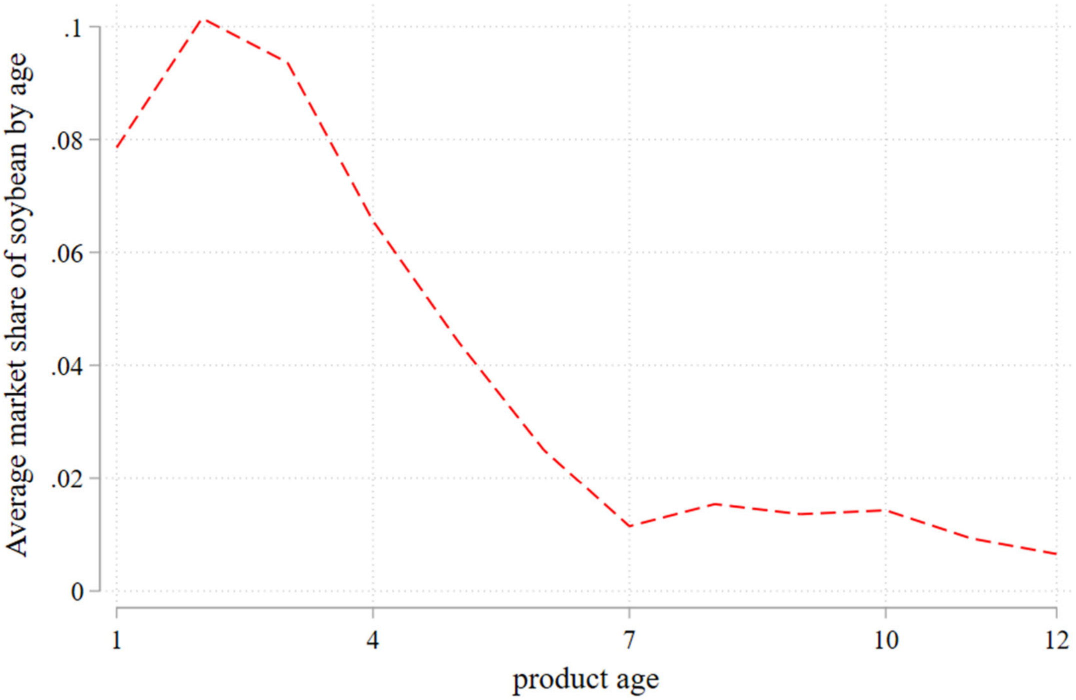
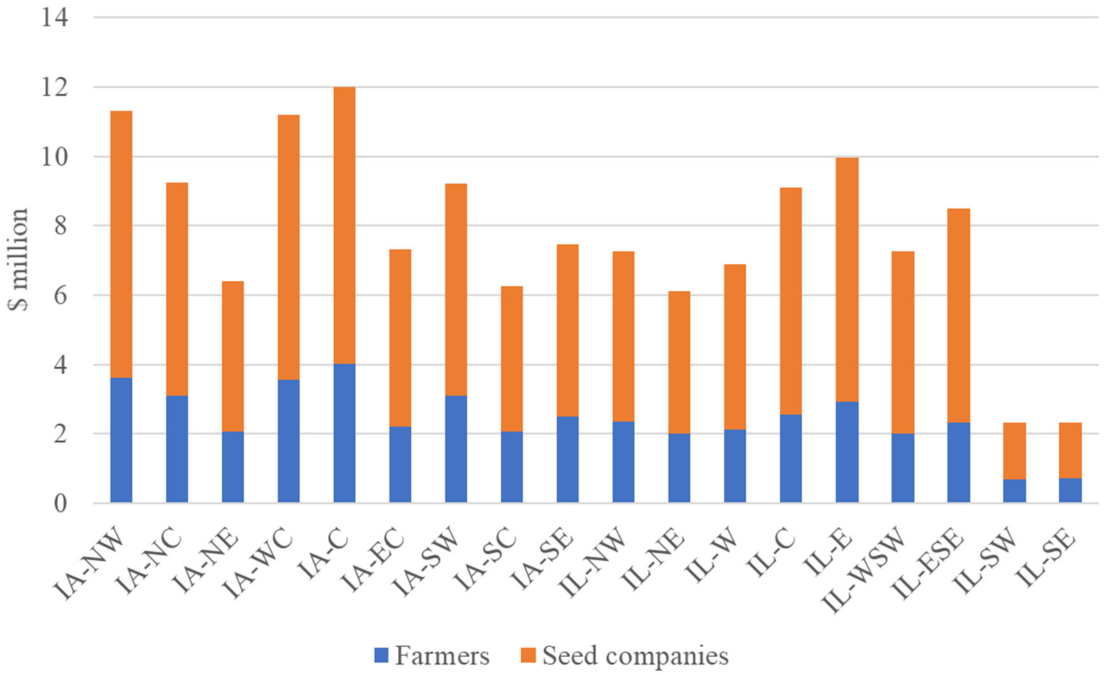


Figure 4. Estimated welfare change due to both SCN-resistant varieties and ISU-SCN information, Iowa and Illinois by CRDs, 2011-2016

Panel (a)



Panel (b)

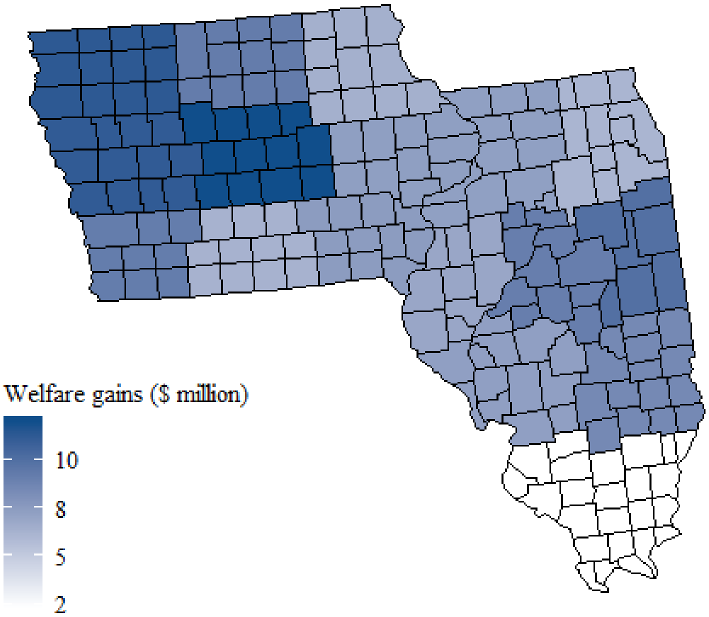


Table 1. Average (nominal) price by trait group, Iowa and Illinois, 2011-2016

	Non resistant		SCN-resistant		Tested by ISU	
Year	Non-GE	GE	Non-GE	GE	Non-GE	GE
2011	38.91	48.68	41.64	49.74	--	48.92
2012	42.76	52.05	45.32	53.76	48.98	53.58
2013	44.91	55.29	45.88	57.67	42.99	58.03
2014	43.97	58.97	46.91	59.23	50.64	58.66
2015	50.72	59.25	51.39	58.76	52.79	59.45
2016	52.80	59.61	51.53	59.15	53.50	59.22
Mean	45.68	55.64	47.11	56.38	49.78	56.31

Source: ISU-SCN data and Kynetec data

Table 2. Number of varieties and market shares, Iowa and Illinois, 2011-2016

Parent company	Brand	Market share	Total number of varieties	Number of SCN-resistant varieties	Number of ISU-SCN tested varieties
AgReliant	LG Seeds	1.46%	80	61	10
	Agreliant-others	1.03%	62	36	1
Dow Agrosciences	Prairie Brand	1.34%	79	58	46
	Mycogen	1.21%	68	50	16
	Agrosci-others	1.57%	103	22	0
Dupont	Pioneer	32.08%	280	107	49
	Dupont-others	1.73%	140	49	20
Monsanto	Asgrow Seed Company	22.09%	210	118	58
	Channel	3.83%	97	37	2
	Kruger Seed	1.63%	74	54	37
	Stone Seed Farms Inc.	1.14%	47	14	2
	Monsanto-others	1.59%	116	44	9
Syngenta	NK Seeds and others	8.72%	135	67	45
Local & Regional Companies	Beck's Hybrids	3.23%	99	35	6
	Stine Seed Company	3.11%	147	94	24
	Growmark / FS	3.09%	115	72	25
	Dyna-Gro	1.71%	73	36	4
	Other-others	9.41%	780	208	63
Total		100%	2,705	1,162	417
of which GT			2,300	1,014	396
of which conventional			405	148	21

Source: ISU-SCN data and Kynetec data

Table 3. Average number of varieties per market (conventional and with GT trait)

	Conventional		GE trait	
Year	Mean	S.D.	Mean	S.D.
2011	7.39	(4.46)	88.37	(22.88)
2012	11.21	(5.31)	85.74	(24.17)
2013	11.31	(6.54)	84.40	(25.24)
2014	15.69	(8.64)	79.91	(21.57)
2015	16.01	(6.65)	85.15	(26.49)
2016	20.73	(8.04)	76.26	(23.70)

Source: ISU-SCN data and Kynetec data

Table 4. Estimated parameters of the demand model

		---------- OLS ----------		---------- 2SLS (IV) ----------	
	$\ln(s_{jm}/s_{0m})$	Coefficient	S.E.	Coefficient	S.E.
α	Price	0.000145	(0.0000987)	-0.0143***	(0.00280)
β_1	SCN-resistant	0.000111	(0.00231)	0.0116**	(0.00501)
β_2	Tested by ISU-SCN	-0.00120	(0.00366)	0.0107	(0.00722)
β_3	Yield top 50%	0.00197	(0.00318)	0.0195***	(0.00677)
β_4	SCN control top 50%	0.00148	(0.00314)	-0.00514	(0.00620)
γ	GT trait	0.00569*	(0.00332)	0.133***	(0.0250)
σ	Nesting corr.	0.999***	(0.000796)	0.947***	(0.0127)
	$1 \leq$ Age ≤ 3	0	(.)	0	(.)
	$4 \leq$ Age ≤ 6	-0.00129	(0.00221)	-0.0201***	(0.00580)
	$7 \leq$ Age ≤ 9	0.00169	(0.00458)	-0.0502***	(0.0136)
	$10 \leq$ Age	-0.00348	(0.00368)	-0.0318***	(0.00996)
	Constant	-0.405***	(0.00730)	-0.129	(0.122)
	Year FE	Y		Y	
	CRD FE	Y		Y	
	Brand FE	Y		Y	
	IVs (6)			Y	
	N	8,984		8,984	

Note: Standard errors in parentheses. Significance: * p<0.10 , ** p<0.05 , *** p<0.01.

Table 5. Hedonic price regression

	Price	Coefficient	Standard error
ϕ_0	GT trait	8.713***	(0.377)
θ_1	SCN resistance	0.698***	(0.259)
θ_2	Tested by ISU-SCN	0.314	(0.383)
θ_3	Yield top 50%	0.664**	(0.332)
θ_4	SCN control top 50%	-0.475	(0.324)
	$1 \leq$ Age ≤ 3	0	(.)
	$4 \leq$ Age ≤ 6	-1.465***	(0.229)
	$7 \leq$ Age ≤ 9	-3.780***	(0.495)
	$10 \leq$ Age	-2.245***	(0.444)
	Constant	37.78***	(0.558)
	Brand FE	Y	
	CRD FE	Y	
	year FE	Y	
	N	8984	

Note: * p<0.1, ** p<0.05, *** p<0.01 .

Table 6. Summary of welfare measures, IA and IL, 2011-2016 ($ million)

		Source of Welfare Gains		
Method	Category	SCN Resistance	ISU-SCN Extension	SCN Resistance & ISU-SCN
Surplus calculation	Total surplus	62.52	65.98	128.50*
	Seed suppliers' revenue change	94.30	65.10	159.40*
	Imputed farmers' net returns	-31.78	0.88	-30.90*
Counterfactual analysis (Keep all)	Total welfare change	61.27	78.80	140.07
	Seed suppliers' revenue change	53.17	43.06	96.23
	Farmers' welfare gains	8.10	35.74	43.84
Counterfactual analysis (Naïve)	Total welfare change	463.17	345.18	808.35
	Seed suppliers' revenue change	192.91	151.74	344.65
	Farmers' welfare gains	270.26	193.44	463.70

Note: For the "Naïve" counterfactual, 417 and 1,162 varieties are dropped from the choice set, out of 2,705 varieties, for the "tested" and "SCN-resistant" varieties, respectively.

SUPPLEMENTARY APPENDIX

Table of contents

Appendix A – Kynetec seed data

The source of seed data used in this study is drawn from a large set of farm-level observations of seed choices by U.S. soybean farmers. In particular, we use the soybean TraitTrak® dataset, a proprietary dataset developed by Kynetec USA, Inc., a unit of a major market research organization that specializes in the collection of agriculture-related survey data (https://www.kynetec.com/tracking-market-trends-agriculture). Kynetec constructs the TraitTrak® data from annual surveys of randomly sampled farmers in the United States. The samples are developed to be representative at the crop reporting district (CRD) level. CRDs are multi-county sub-state regions identified by USDA's National Agricultural Statistics Service. **Table X1** reports some descriptive statistics concerning these data. The data available to us covers the 21-year period 1996-2016. We utilize the entire dataset to characterize some properties of seeds at the variety level (e.g., commercial age), although the econometric analysis is performed with data for Iowa and Illinois only, and for the period 2011-2016. Nationwide, the data are based on responses from an average of 3,613 soybean farmers per year over the entire 1996-2016 period. For the two states and the six years that are the focus of our analysis—Iowa and Illinois, 2011-2016—there are an average of 901 soybean farmers per year.

Dropped Observations. In the subset of the soybean TraitTrak® data for Iowa and Illinois during 2011-2016, we drop 13 observations (out of 16,386 soybean seed purchases) in which the planted seeds were sold back to seed firms. Also, we drop four observations whose brand variable was coded as "unidentified" or "unspecified."

Size of Market and Market Shares. We define the potential size of a market by using acres planted to corn or soybean. To calculate the potential size of market, the projected acres from seed purchases are required. For that purpose, we draw on corn TraitTrak® data as well as soybean TraitTrak® data. The average size of a "market," on our study, is about 2.8 million acres, 59% of which is taken by corn (the outside option in our model). Given a pair of CRD and year, each variety's planted acres are divided by the market size (total acres planted to either corn or soybeans), which returns individual product's market share.

Appendix B – Further description of the ISU-SCN data

As shown in **Figure 1** in the text, the ISU-SCN dataset is a key component of our analysis. This data source is developed by the *Laboratory of Dr. Gregory L. Tylka* at Iowa State University. In particular, the ISU-SCN dataset contains detailed growth records and seed characteristics of tested varieties. In this appendix, we provide a more detailed description of the ISU-SCN data and how we convert them to a tractable form for our analysis. **Table X2** reports descriptive statistics of ISU-SCN data.

As this data set is collected from experimental field trials, susceptible varieties are included in every region and time of tests as a control group. These susceptible varieties are popular varieties grown in Iowa that lack the SCN-resistance trait. Metrics of interest concern the yield performance and end-of-season SCN population density (pest pressure). On average, the SCN-resistant group has a higher yield rate and lower SCN population density than the susceptible group. **Figure X1** illustrates the distribution of test locations across Iowa by reporting the number of tested varieties, by counties, over the entire period. **Figure X2** shows the correlation between end-of-season SCN population density and yield. The fitted lines tell us that the group of SCN-resistant varieties is more resilient to an environment with high SCN population density (high pest pressure), and this improved resilience is translated into higher yield.

B.1. Constructing Performance Scores

As reported in **Table X2**, SCN density varies a lot, which is commonly attributed to the soil conditions of the test site. Also, since each variety is not tested in all sites, a direct comparison of yield or SCN population density between tested varieties is not appropriate. To control for this regional effect on performances, we generate performance scores of individual varieties relative to the control group. To be specific, given a test region r, the performance score of variety j is

$$score(X)_{rj} = \frac{X_{rj}}{\bar{X}_{r0}} \text{ for } X \in \{yield, SCN density\}$$

where $\bar{X}_{r0}$ is the average performance of the control group in region *r*, and region *r* is one of the nine districts of Iowa ($r \in \{NW, NC, NE, WC, C, EC, SW, SC, SE\}$). To obtain the relative performance of a variety tested at multiple locations, we aggregate regional scores to an individual variety score.[1] The aggregate performance score of variety j is

$$score(X)_j = E[score(X)_{rj}].$$

For the yield score, a higher value indicates superiority, while the lower rating is preferred for the SCN population density score. As stated in the paper, we set the 50 percentile as a threshold for a better performing group in each annual report (i.e., the better performing group is decided by comparison within varieties tested in the same year).

B.2. Varieties Tested Multiple Times

Some varieties were tested several times. In this case, it is natural to assume that farmers refer to the latest information when they purchase. For example, farmers who purchases seeds in 2014 could see a variety that was tested in 2012 and 2013. Then, they may refer 2013 information rather than 2012. Thus, in our analysis, the extension information variables are defined by the latest test scores. **Table X3** summarizes the number of varieties with the number of tested varieties.

---

[1] For instance, in 2015, a south-central Iowa experimental field was significantly damaged by sudden death syndrome, which influences measurement of yield performances. In this case, other regional scores are considered more carefully. Thus, simply using specific regional scores cannot properly reflect how farmers receive the information.

Appendix C – Robustness checks

In Appendix C, we report several results with different empirical specifications about the extension information and the product age variable for robustness checks.

C.1. Different Specifications for Extension Information

In our main model, we use the 50th percentile as a threshold for the high performing group. However, the threshold can be questioned because farmers can use a different level of threshold. Thus, it is important to do robustness checks using differently defined information variables.

Model A1. Using the top 25th percentile for the threshold of a high performing group

First, we use the 25th percentile for the alternative threshold for a high-performance group. In this case, we only need to replace $\mathbb{I}_{jt[m]}^{\text{y-top50}}$ and $\mathbb{I}_{jt[m]}^{\text{s-top50}}$ with $\mathbb{I}_{jt[m]}^{\text{y-top25}}$ and $\mathbb{I}_{jt[m]}^{\text{s-top25}}$ in the demand estimation and hedonic price regression (respectively, equation (12) and (13) in the main text). **Table X4** and **Table X5** respectively report the demand estimation results and hedonic price regression. **Table X6** shows the welfare measure of this specification.

Model A2. Using the top 10th percentile for the threshold of a high performing group

Similar to Model A1, we also consider the 10th percentile as an alternative threshold for the high-performance group. **Table X7** and **Table X8** respectively present the demand estimation results and hedonic price regression. **Table X9** reveals the welfare measure of this specification.

Summarizing the results of these two models, although the different definition of the information factor affects interpretation of WTPs and hedonic price regression, we see that all specifications reported here provide considerably similar welfare findings as our main model. This bolsters the robustness of our main model.

C.2. Different Specifications for Age Variables

In our main model, we use four age-related dummy variables to control for the varieties' product life cycle. For the robustness of our results, especially for the product age, here we try two additional specifications for the age variable.

Model A3. Using more product age fixed effects to differentiate every age

As an alternative specification for the product age, we can add more specific age dummies. In **Model A3**, we introduce eleven product age dummies; ten individual age fixed effects and an eleventh dummy which codes all products older than 10 years. **Table X10**, **Table X11**, and **Table X12** summarize the results.

Model A4. Use log of age instead of product age fixed effects

Instead of the age fixed effects, product age can be directly used in the regression equations. To be specific, log of the product age is used to account for the product life cycle. **Table X13**, **Table X14**, and **Table X15** summarize the results.

Both **Model A3** and **Model A4** return quite close results to our main model. Therefore, the specification in the main text appear to be robust.

Appendix D – Supplementary figures and tables

Figure X1. Number of observations in ISU SCN data (1997-2015)

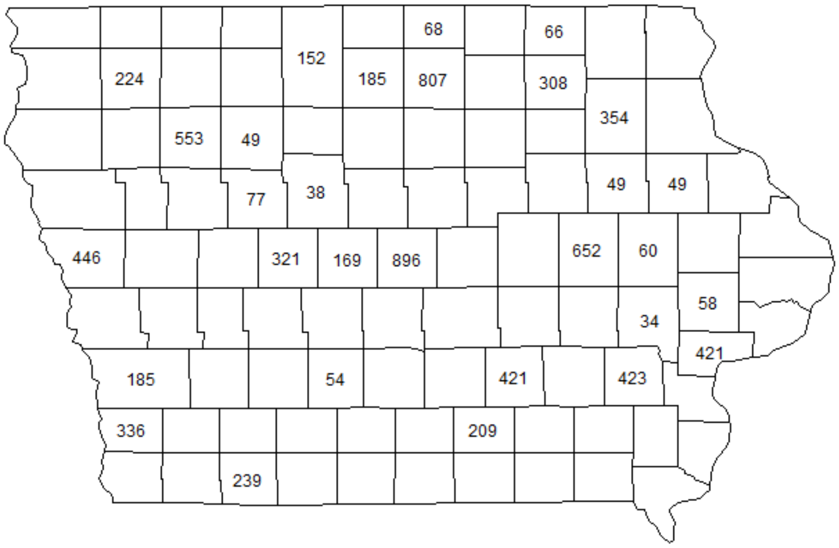
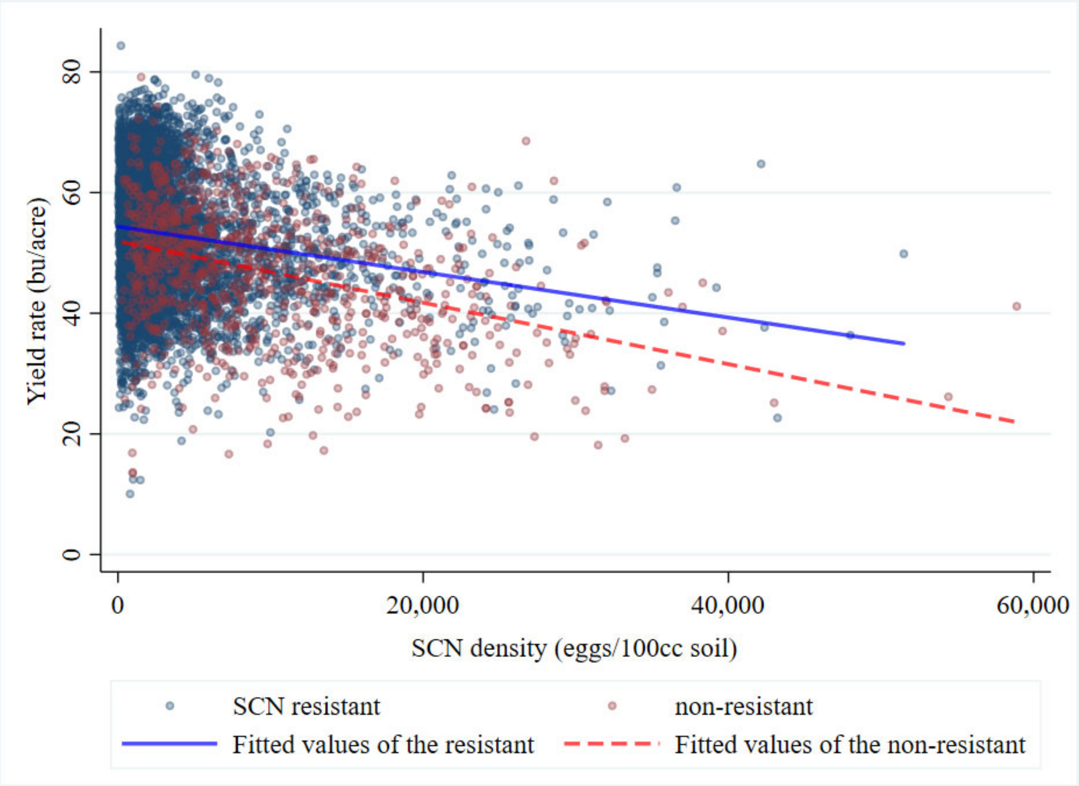


Figure X2. Relationship between SCN density and yield rate



Source: ISU-SCN data

Table X1. Descriptive summary of Kynetec soybean seed data, 1996-2016

	U.S. 1996-2016	IA and IL 1996-2016	IA and IL 2011-2016
No. of states represented	29	2	2
No. of CRDs represented	179	18	18
No. of farms per year	3,613	1,003	901
No. of transactions per farm	2.8	3.1	3.0
No. of varieties sold per year	2,146	867	767
No. of brands per year	189	92.7	69.5
Soybean acres per market			1,161,817
Corn acres per market			1,649,112

Table X2. Summary statistics of ISU-SCN data, 1997-2015

		SCN-resistant	SCN-susceptible (Control group)
No. of observations		6,535	863
No. of varieties		1,798	106
No. of Glyphosate-tolerant varieties		1,391	43
Yield (bu/acre)	Mean	52.05	46.70
	Standard Dev.	9.08	10.48
SCN density (eggs/100cc soil)	Mean	3,039.5	9,220.2
	Standard Dev.	4,294.7	7,740.6

Table X3. Number of varieties tested multiple times

Frequency of being tested	SCN-resistant	SCN-susceptible
1	1,405	52
2	278	17
3	78	24
4	23	9
5	7	1
6	2	2
7	2	0
8	1	0
9	2	1
Total	1,798	106

Table X4. Estimated parameters of seed demand with top 25% performance group

		---------- OLS ----------		---------- 2SLS (IV) ----------	
	$\ln(s_{jm}/s_{0m})$	Coefficient	S.E.	Coefficient	S.E.
α	Price	0.000143	(0.0000987)	-0.0148***	(0.00290)
β_1	SCN-resistant	0.000126	(0.00231)	0.0120**	(0.00514)
β_2	Tested by ISU-SCN	-0.0000421	(0.00276)	0.0159**	(0.00630)
β_3	Yield top 25%	0.00362	(0.00345)	0.0206***	(0.00727)
β_4	SCN control top 25%	-0.00103	(0.00347)	-0.00823	(0.00711)
γ	GT trait	0.00570*	(0.00332)	0.137***	(0.0259)
σ	Nesting corr.	0.999***	(0.000795)	0.946***	(0.0131)
	$1 \leq$ Age ≤ 3	0	(.)	0	(.)
	$4 \leq$ Age ≤ 6	-0.00125	(0.00222)	-0.0207***	(0.00596)
	$7 \leq$ Age ≤ 9	0.00174	(0.00457)	-0.0520***	(0.0140)
	$10 \leq$ Age	-0.00344	(0.00368)	-0.0330***	(0.0103)
	Constant	-0.404***	(0.00722)	-0.137	(0.124)
	Year FE	Y		Y	
	CRD FE	Y		Y	
	Brand FE	Y		Y	
	IVs (6)			Y	
	N	8,984		8,984	

Note: Standard errors in parentheses. Significance: * p<0.10 , ** p<0.05 , *** p<0.01.

Table X5. Hedonic price regression of Model A1

	Price	Coefficient	Standard error
ϕ_0	GT trait	8.706***	(0.377)
θ_1	SCN-resistant	0.694***	(0.259)
θ_2	Tested by ISU-SCN	0.352	(0.302)
θ_3	Yield top 25%	0.658*	(0.346)
θ_4	SCN control top 25%	-0.390	(0.367)
	$1 \leq$ Age ≤ 3	0	(.)
	$4 \leq$ Age ≤ 6	-1.460***	(0.229)
	$7 \leq$ Age ≤ 9	-3.784***	(0.494)
	$10 \leq$ Age	-2.251***	(0.444)
	Constant	37.79***	(0.558)
	Brand FE	Y	
	CRD FE	Y	
	year FE	Y	
	N	8984	

* p<0.1 , ** p<0.05 , *** p<0.01

Table X6. Summary of welfare measures of Model A1

		Source of Welfare Gains		
Method	Category	SCN Resistance	ISU-SCN Extension	SCN Resistance & ISU-SCN
Surplus calculation	Total surplus	61.95	66.08	128.03
	Seed suppliers' revenue change	94.30	67.02	161.32
	Imputed farmers' net returns	-32.35	-0.94	-33.29
Counterfactual analysis (Keep all)	Total welfare change	58.31	78.94	137.25
	Seed suppliers' revenue change	54.58	43.93	98.51
	Farmers' welfare gains	3.73	35.01	38.74
Counterfactual analysis (Naïve)	Total welfare change	461.33	341.16	802.49
	Seed suppliers' revenue change	196.70	151.71	348.41
	Farmers' welfare gains	264.63	189.45	454.08

Table X7. Estimated parameters of seed demand with top 10% performance group

		---------- OLS ----------		---------- 2SLS (IV) ----------	
	$\ln(s_{jm}/s_{0m})$	Coefficient	S.E.	Coefficient	S.E.
α	Price	0.000143	(0.0000988)	-0.0145***	(0.00283)
β_1	SCN-resistant	0.000121	(0.00231)	0.0117**	(0.00506)
β_2	Tested by ISU-SCN	0.000697	(0.00252)	0.0182***	(0.00588)
β_3	Yield top 10%	0.00216	(0.00481)	0.0262***	(0.00967)
β_4	SCN control top 10%	-0.00236	(0.00508)	-0.0307***	(0.0113)
γ	GT trait	0.00572*	(0.00332)	0.134***	(0.0253)
σ	Nesting corr.	0.999***	(0.000796)	0.947***	(0.0128)
	$1 \leq$ Age ≤ 3	0	(.)	0	(.)
	$4 \leq$ Age ≤ 6	-0.00132	(0.00221)	-0.0206***	(0.00587)
	$7 \leq$ Age ≤ 9	0.00181	(0.00458)	-0.0505***	(0.0137)
	$10 \leq$ Age	-0.00347	(0.00368)	-0.0325***	(0.0101)
	Constant	-0.404***	(0.00723)	-0.141	(0.122)
	Year FE	Y		Y	
	CRD FE	Y		Y	
	Brand FE	Y		Y	
	IVs (6)			Y	
	N	8,984		8,984	

Note: Standard errors in parentheses. Significance: * p<0.10 , ** p<0.05 , *** p<0.01.

Table X8. Hedonic price regression of Model A2

	Price	Coefficient	Standard error
ϕ_0	GT trait	8.704***	(0.377)
θ_1	SCN-resistant	0.695***	(0.259)
θ_2	Tested by ISU-SCN	0.470*	(0.277)
θ_3	Yield top 10%	0.720	(0.444)
θ_4	SCN control top 10%	-1.526***	(0.555)
	1 ≤ Age ≤ 3	0	(.)
	4 ≤ Age ≤ 6	-1.470***	(0.229)
	7 ≤ Age ≤ 9	-3.774***	(0.494)
	10 ≤ Age	-2.253***	(0.445)
	Constant	37.80***	(0.558)
	Brand FE	Y	
	CRD FE	Y	
	year FE	Y	
	N	8984	

* p<0.1 , ** p<0.05 , *** p<0.01

Table X9. Summary of welfare measures of Model A2

		Source of Welfare Gains		
Method	Category	SCN Resistance	ISU-SCN Extension	SCN Resistance & ISU-SCN
Surplus calculation	Total surplus	62.05	68.21	130.26
	Seed suppliers' revenue change	94.30	58.76	153.06
	Imputed farmers' net returns	-32.35	9.45	-22.8
Counterfactual analysis (Keep all)	Total welfare change	58.31	78.94	137.25
	Seed suppliers' revenue change	54.58	44.70	99.39
	Farmers' welfare gains	3.73	35.44	39.03
Counterfactual analysis (Naïve)	Total welfare change	456.34	342.57	798.91
	Seed suppliers' revenue change	192.19	152.97	345.16
	Farmers' welfare gains	264.15	189.60	453.75

Table X10. Estimated parameters of the demand model with 11 product age dummies

		---------- OLS ----------		---------- 2SLS (IV) ----------	
	$\ln(s_{jm}/s_{0m})$	Coefficient	S.E.	Coefficient	S.E.
α	Price	0.000150	(0.0000989)	-0.0145***	(0.00278)
β_1	SCN-resistant	0.000194	(0.00231)	0.0117**	(0.00501)
β_2	Tested by ISU-SCN	-0.00161	(0.00368)	0.0116	(0.00731)
β_3	Yield top 50%	0.00209	(0.00318)	0.0183***	(0.00672)
β_4	SCN control top 50%	0.00166	(0.00314)	-0.00540	(0.00623)
γ	GT trait	0.00557*	(0.00334)	0.134***	(0.0250)
σ	Nesting corr.	0.999***	(0.000798)	0.950***	(0.0128)
	Age 1	0	(.)	0	(.)
	Age 2	-0.000152	(0.00241)	-0.000791	(0.00552)
	Age 3	0.00280	(0.00281)	-0.00519	(0.00712)
	Age 4	-0.00141	(0.00329)	-0.0189**	(0.00832)
	Age 5	-0.000909	(0.00370)	-0.0277***	(0.00972)
	Age 6	0.00272	(0.00471)	-0.0219**	(0.0109)
	Age 7	0.00234	(0.00659)	-0.0514***	(0.0163)
	Age 8	-0.000340	(0.00866)	-0.0382*	(0.0215)
	Age 9	0.00522	(0.00894)	-0.0689***	(0.0231)
	Age 10	-0.00502	(0.0108)	-0.0752***	(0.0256)
	Age ≥ 11	-0.00254	(0.00405)	-0.0295***	(0.0112)
	Constant	-0.405***	(0.00730)	-0.129	(0.122)
	Year FE	Y		Y	
	CRD FE	Y		Y	
	Brand FE	Y		Y	
	IVs (6)			Y	
	N	8,984		8,984	

Note: Standard errors in parentheses. Significance: * p<0.10 , ** p<0.05 , *** p<0.01.

Table X11. Hedonic price regression of Model A3

	Price	Coefficient	Standard error
ϕ_0	GT trait	8.763***	(0.377)
θ_1	SCN-resistant	0.684***	(0.258)
θ_2	Tested by ISU-SCN	0.541	(0.386)
θ_3	Yield top 50%	0.580*	(0.332)
θ_4	SCN control top 50%	-0.527	(0.324)
	Age 1	0	(.)
	Age 2	-0.597**	(0.281)
	Age 3	-1.364***	(0.306)
	Age 4	-1.895***	(0.340)
	Age 5	-2.308***	(0.402)
	Age 6	-2.187***	(0.523)
	Age 7	-3.994***	(0.708)
	Age 8	-3.468***	(1.016)
	Age 9	-5.928***	(0.899)
	Age 10	-5.413***	(1.042)
	Age ≥ 11	-2.515***	(0.499)
	Constant	38.07***	(0.566)
	Brand FE	Y	
	CRD FE	Y	
	year FE	Y	
	N	8984	

* p<0.1 , ** p<0.05 , *** p<0.01

Table X12. Summary of welfare measures of Model A3

		Source of Welfare Gains		
Method	Category	SCN Resistance	ISU-SCN Extension	SCN Resistance & ISU-SCN
Surplus calculation	Total surplus	62.15	65.21	127.35
	Seed suppliers' revenue change	94.30	65.10	159.40
	Imputed farmers' net returns	-32.16	0.11	-32.05
Counterfactual analysis (Keep all)	Total welfare change	61.65	75.37	137.02
	Seed suppliers' revenue change	52.90	45.23	98.13
	Farmers' welfare gains	8.75	30.14	38.89
Counterfactual analysis (Naïve)	Total welfare change	439.31	329.33	768.64
	Seed suppliers' revenue change	186.27	148.00	334.27
	Farmers' welfare gains	253.04*	181.33	434.37

Table X13. Estimated parameters of the demand model with log of product age

		---------- OLS ----------		---------- 2SLS (IV) ----------	
	$\ln(s_{jm}/s_{0m})$	Coefficient	S.E.	Coefficient	S.E.
α	Price	0.000147	(0.0000988)	-0.0148***	(0.00291)
β_1	SCN-resistant	0.000472	(0.00228)	0.0124**	(0.00503)
β_2	Tested by ISU-SCN	-0.00123	(0.00364)	0.0121	(0.00737)
β_3	Yield top 50%	0.00207	(0.00317)	0.0187***	(0.00686)
β_4	SCN control top 50%	0.00150	(0.00314)	-0.00663	(0.00639)
γ	GT trait	0.00571*	(0.00332)	0.137***	(0.0262)
σ	Nesting corr.	0.999***	(0.000797)	0.948***	(0.0130)
	log(Age)	-0.000314	(0.00120)	-0.0155***	(0.00450)
	Constant	-0.404***	(0.00727)	-0.127	(0.126)
	Year FE	Y		Y	
	CRD FE	Y		Y	
	Brand FE	Y		Y	
	IVs (6)			Y	
	N	8,984		8,984	

Note: Standard errors in parentheses. Significance: * p<0.10 , ** p<0.05 , *** p<0.01.

Table X14. Hedonic price regression of Model A4

	Price	Coefficient	Standard error
ϕ_0	GT trait	8.778***	(0.378)
θ_1	SCN-resistant	0.643**	(0.253)
θ_2	Tested by ISU-SCN	0.455	(0.382)
θ_3	Yield top 50%	0.582*	(0.332)
θ_4	SCN control top 50%	-0.570*	(0.324)
	log(Age)	-1.296***	(0.139)
	Constant	38.11***	(0.563)
	Brand FE	Y	
	CRD FE	Y	
	year FE	Y	
	N	8984	

* p<0.1 , ** p<0.05 , *** p<0.01

Table X15. Summary of welfare measures of Model A4

		Source of Welfare Gains		
Method	Category	SCN Resistance	ISU-SCN Extension	SCN Resistance & ISU-SCN
Surplus calculation	Total surplus	64.47	63.99	128.46
	Seed suppliers' revenue change	94.3	65.1	159.4
	Imputed farmers' net returns	-29.83	-1.11	-30.94
Counterfactual analysis (Keep all)	Total welfare change	66.26	76.52	142.78
	Seed suppliers' revenue change	52.61	43.61	96.22
	Farmers' welfare gains	13.65	32.91	46.56
Counterfactual analysis (Naïve)	Total welfare change	448.73	336.09	784.82
	Seed suppliers' revenue change	288.52	152.02	440.54
	Farmers' welfare gains	160.21	184.07	344.28